THE UNIVERSITY
of EDINBURGH

SMAD2/3 POTENTIATE CELL IDENTITY CONVERSIONS WITH MASTER TRANSCRIPTION FACTORS

Tyson Ruetz

Thesis presented for the degree of Doctor of Philosophy

MRC Centre for Regenerative Medicine

University of Edinburgh

2015

Declaration

I declare that the work presented in this thesis is my own, unless otherwise stated, and has not been submitted for any other degree or professional qualification.

_____

Acknowledgments

I am grateful to have worked on a PhD in Kei's lab. He was an attentive supervisor, who pushed me to get the most out of my work. He is an excellent scientist and mentor, and I have come away with a great appreciation for the fine details of molecular biology and experimental design. It has been a very enjoyable 4 years, and I am sure we will stay in contact for years to come. Further, a massive thank you to my co-supervisors Sally and Tilo, who have provided great feedback and were very supportive and encouraging over the past 4 years.

This project was the results of collaboration amongst four labs. I had the pleasure of working with Thomas Graf from the Centre for Genomic Regulation, Barcelona, Malin Parmar from Lund University, and Sten Linnarsson from the Karolinska Intitute. A special thank you to Ulrich Pfisterer, Bruno DiStefano and Anna Johnsson for all their help and beautiful contributions to the work.

I have had the privilege of working with many great scientists, who have shaped me and made my time in Edinburgh a beautiful experience. My first day in Edinburgh I went for drinks with the Kaji lab, who had been incredibly generous and supportive during my move across the Atlantic to Edinburgh, and continue to be wonderful friends. Thank you to all the lab members, past and present, which have provided great feedback and support over the years. A brief thank you to the many characters of the lab including James for being a wonderful friend and training me in the lab, Sara for being a lovely bench neighbour, Eleni for putting up with men/children all day, Luca for being a great lab manager and motivating us with your passion and determination, Dan for the amazing banter and being a true friend, Mig(w)uel for all your Miguelisms, Nic for the daily inappropriate comments, Kumiko for leaving in spectacular fashion, JamesA for his ability to work with me while temporarily ignoring Luca's daily pestering and Sergio

for mentoring me and guiding the project. It has been a great ride, with many laughs and experiences together. I will forever miss the coffee breaks and banter in the lab! I am sure we will all stay in touch and I will enjoy watching you all develop and succeed wherever life takes you.

To the many friends and colleagues at the institute, thank you for making my time so wonderful. Thank you to Kousa for broadcasting our tennis results 'live' to the institute (WHEN you won) and for turning up at half 10 everyday, making the rest of us feel good about our work ethic, Sabine (and Matthias) for always bringing a sense of adventure, the Krank group for entertaining us with your craziness, Raph for amazing first year stories, Zaner for introducing us to squash, the Medvinsky crew for sharing stories from your lab, Filip and Eleni for playing host to many great parties and late nights, and everyone else who made my time in the centre such an enjoyable experience.

To my flatmate and fellow PhD student, Harsh, thank you for the great friendship and many laughs over the years! We have come a long way from those first journal club meetings and long nights in Hydra! I would like to say that things have only gotten better, but lets be honest, Hydra was the pinnacle of our PhD! Thank you for the many great experiences together, and I look forward to a lifetime of friendship.

Moving to a new country, leaving friends and family behind, was a challenging venture that was eased by my partner in crime and loving wife, Laurel. Throughout the move and acclimation to Edinburgh, we developed a very strong bond that I believe is now the foundation of our relationship. From the beaches of Greece and spectacular hikes throughout the Highlands, to our evening strolls in majestic Edinburgh, I will always cherish our time in Scotland. Thank you for continually being a positive reinforcement in my life.

I would like to dedicate this work to my Brother, Mom and Dad, who have been the greatest inspiration in my life. To my parents, you have provided constant encouragement and inspiration. I always try to emulate your work ethic and genuine appreciation in life. Thank you for guiding me on this beautiful journey. Adam has been my best friend and a genuine role model throughout life, teaching me that we can do

anything we dedicate ourselves to. Also, to my Grandparents, Harold and Mae Ruetz, and Doreen Badgero, you are forever my guiding light and reason I strive to make the most of life and enjoy the ride.

Abstract

The exogenous expression of master transcription factors (TFs) to drive cell identity changes is an exciting and powerful approach to cell and tissue engineering. Yet, the generation of desired cell types is often plagued by inefficiency and inability to produce mature cell types. Through investigations of the molecular mechanisms of induced pluripotent stem cell (iPSC) generation, I discovered that expression of constitutively active *Smad2/3* (*Smad2CA/3CA*), together with the Yamanaka factors, could dramatically improve the efficiency of reprogramming. Mechanistically, SMAD3 interacted with both co-activators and reprogramming factors, bridging their interaction during reprogramming. Because SMAD2/3 interact with a multitude of master TFs in different cell types, I tested the conversions of B cells to macrophages, myoblasts to adipocytes, and human fibroblasts to neurons. Remarkably, each conversion system was markedly enhanced when the master TFs were co-expressed with *Smad3CA*. These results revealed the existence of shared molecular mechanisms underlying diverse TF-mediated cellular conversions, and demonstrated SMAD2/3 as a widely applicable co-factor that potentiates the generation of diverse cell types with profound efficiency and maturity.

Lay Summary

Stem cells are a cell type capable of generating any tissue and cell of the body. Recent advances in cell and molecular biology have enabled stem cells to be used to create cells and tissue types that could one day be utilized for regenerative therapies. However, there are a number of challenges facing the field before clinical trials take place; chiefly, the potential for immune rejection upon transplantation of stem cell derived tissue is a major roadblock. A technology exists to produce stem cells from a patient's own cells, which could reduce the immune response upon transplantation. This thesis describes a new technology for the production of stem cells from skin cells with improved efficiency. Furthermore, the technology can facilitate generation of other clinically relevant cells types that could also be used for regenerative therapies.

Table of Contents

List of Tables and Figures

1: Introduction

1.1 Changing cell identity

1.1.1 Somatic cell nuclear transfer

Developmental biologists have for years probed the subject of how specialized, or differentiated, cells emerge from unspecialized embryonic cells. For the first half the 20th century, differentiation hypotheses were centred on the nucleus, suggesting its physical and positional properties were manipulated during cell fate specification. Such hypotheses often suggested that during development, genes or chromosomes were physically sorted or permanently altered, either by deletion or mutation, to give rise to somatic cell types. Yet many scientists refused to accept that coordinated deletions and mutations in genes occurred in such fine detail, given the infrequent rates of mutation they had observed (King and Briggs, 1955). Thus King and Briggs set out to investigate if differentiating cell nuclei contained all the genetic information to that of an embryonic cell. In the early 1950's, they developed a technique to transfer the nucleus from one cell to another enucleated cell, allowing them to ask an important question: can a nucleus of a certain cell type change its function when placed in another cell. The first experiments involved transferring early embryonic cells from the frog species *Rana pipiens*, into enucleated eggs of the same species and then observing their capacity to develop. Indeed the recipient egg underwent normal development, suggesting that there were at least no lasting effects on nuclei during the pre-gastrulation stage of development. They went on to test transfers taking nuclei from different developmental stage cells into enucleated eggs, demonstrating that in most cases, a fraction of the cells had the capacity to undergo normal development up to the blustula or gastrula stage, upon which development arrested. King and Briggs then erroneously concluded that the nuclei of later stage

developed embryos were restricted as a result of definite nuclear changes during differentiation. However, their seminal work influenced another developmental biologist, John Gurdon, to further pursue the question of whether a committed adult somatic cell nucleus could give rise to adult organisms if presented with optimal conditions. Indeed Gurdon first proved that, through optimizing the technique, a small proportion (3%) of committed tadpole endoderm cell nuclei could be implanted in enucleated eggs which then developed to the tadpole stage of development (Gurdon, 1960). He then followed that publication by demonstrating that a portion of those eggs could fully develop into fertile adults that appeared completely normal (Gurdon, 1962). Thus, Gurdon's studies provided the first evidence that committed somatic cell nuclei could be reprogrammed to a pluripotent state. Those results implied that committed tissue specific cells, when presented with the right signalling factors and environment, contain all the required genetic information to become any other cell type. Further proof came from studies showing that adult cells from mammals, such as sheep, could also give rise to viable adult clones (Wilmut et al., 1997). However skeptics rightfully questioned the potential that contaminating cells, ones less differentiated, may have explained Gurdon and Wilmut's results. The debate was finally ended in 2002, when a group revealed that transferred B-cell nuclei, with traceable genetic rearrangements in the immunoglobulin alleles, were efficiently reprogrammed to pluripotency by nuclear transfer and produced viable adult mice (Hochedlinger and Jaenisch, 2002).

1.1.2 Forced cell identity changes

The idea that cytoplasmic factors may contribute to cell identity sprung a new field of biological studies. Some of the first direct evidence that cytoplasmic factors direct gene expression profiles came from the study of cell fusions. In such experiments, cells are grown in close contact, with polyethylene glycol (PEG) added to the media to promote cell agglutination and membrane fusion. The result is a multinucleated cell that either continues to divide and merge nuclei, or one in which the cell cycle arrests and the cell

remains as a multinucleated cell, termed a heterokaryon (Blau et al., 1983). One such experiment fused human amniocytes with mouse muscle cells and then probed the response of the human nuclei. Through detection of species-specific muscle proteins, the authors demonstrated that within 24 hours, the human amniocytes began to express muscle genes (Blau et al., 1983). Other researchers began searching for the factors that may drive gene expression profile changes, such as those observed in the cell fusion studies. A crucial experiment emerged in studies of antenna development in Drosophila fruit flies, whereby ectopic expression of a single gene, *Antennapedia,* at distinct developmental stages could alter the body plan of the fly, producing legs where antennae normally developed (Schneuwly et al., 1987). In the same year, another group was studying the process by which addition of 5-azacytidine was able to convert fibroblasts to the highly similar myoblasts cell type *in vitro* (Davis et al., 1987). Their work lead to the identification of 3 genes that appeared uniquely and highly expressed only in myoblasts. When they over expressed one such gene, *MyoD*, they found it was sufficient to convert fibroblasts to myoblasts at roughly 50% efficiency (Davis et al., 1987). This perhaps represented the birth of the field of forced cell identity changes by over expression of transcription factors.

1.1.3 Induced pluripotent stem cells

In the early 2000's, the question still remained as to what factors drove the identity change of somatic nuclei to a pluripotent state upon nuclear transfer to oocytes. In 2006, Kazutoshi Takahashi and Shinya Yamanaka combed the literature and produced a list of pluripotency-associated genes, which promoted self-renewal and proliferation in stem cells or tumours (Takahashi and Yamanaka, 2006). They hypothesized that over-expression of these genes could reprogram fibroblasts to pluripotency. By retroviral over-expression of the 24 genes in mouse embryonic fibroblasts, they were able to generate induced pluripotent stem cells (iPSCs); it was an astounding breakthrough. They then performed sequential depletion of the 24 candidates, narrowing the list to just

4 genes; Oct3/4 (O), Sox2 (S), Klf4 (K) and c-Myc (M), as the minimum required factors to induce pluripotency in fibroblasts (Takahashi and Yamanaka, 2006). They subsequently showed that adult tail-tip mouse fibroblasts could also be reprogrammed with the same 4 factors (Takahashi and Yamanaka, 2006). The resultant iPSCs had similar expression profiles to embryonic stem (ES) cells, could be propagated in the same media, had demethylation of pluripotency loci promoters, and could form teratomas; a tumor which forms upon subcutaneous injection of pluripotent cells into adult mice and contains differentiated cells of the three germ layers (Stevens and Little, 1954; Takahashi et al., 2007; Takahashi and Yamanaka, 2006). As a final proof of pluripotency, the mouse iPSCs were injected into blastocysts and were able integrate and contribute to the three germ layers of the embryo. Perhaps the only caveat of these findings was that the chimeric embryos could not develop to full term adult mice (Takahashi and Yamanaka, 2006). However, in less than a year, the Yamanaka lab and others successfully produced iPSCs that gave rise to both chimeric and germline competent adult mice by targeting a selection cassette to *Nanog* or *Oct3/4*, pluripotency associated gene loci, enabling more stringent selection conditions (Okita et al., 2007; Wernig et al., 2007). Follow-up studies also demonstrated the ability to reprogram human fibroblasts to pluripotency (Takahashi et al., 2007). Thus a powerful and reproducible, albeit inefficient, technique had emerged to create genuine pluripotent cells from adult somatic cells through the exogenous expression of just 4 transcription factors.

1.2 Dissecting the process of reprogramming

While the production if iPSCs was highly reproducible among many labs across the world, the process itself remained wildly inefficient (~0.01%), and therefore hampered our understanding of the molecular details of reprogramming. Most of the initial iPSC papers utilized retroviruses to transduce cells with the 4 factors, with significant drawbacks of such systems. Firstly, retroviruses can only infect actively proliferating cells, which may partially explain the low efficiency of reprogramming observed (Cepko and Pear, 2001; Takahashi and Yamanaka, 2006)). Furthermore, the initial studies

expressed the 4 factors on individual cassettes, which would result in heterogeneous expression, and may have contributed to producing the partially reprogrammed colonies observed. The first attempts to address such problems used lentiviral vectors, which can infect non-dividing cells, and expressed the 4 factors off a single cassette (Carey et al., 2009; Shao et al., 2009; Sommer et al., 2009). Such polycistronic cassettes enable more uniform production of exogenous factors by linking multiple genes with viral 2A sequences. The 2A peptides originate from viral polypeptide sequences and contain a well-conserved 18 amino acid sequence. They are found between viral genes and mediate protein cleavage during translation. The 2A region initiates a ribosomal 'skip', whereby a peptide bond is not formed between a proline and glycine within the 2A peptides, thus releasing the upstream translated protein from the ribosome while allowing continued translation of the downstream transcript (Donnelly et al., 2001; Szymczak et al., 2004). While the efficiency of reprogramming increased with the lentiviral delivery and 2A polypeptide approaches, there still remained concerns about using viruses to produce iPSCs, centralized around concerns of genomic insertion mutagenesis as well as aberrant gene activation by viral enhancers in reprogrammed cells. Thus, groups sought to reprogram cells with non-viral strategies. The first robust, virus-free, reprogramming systems utilized the *piggyBac* transposon based plasmid delivery system containing the 4 factors flanked by the transposase inverted *piggyBac* repeats (Kaji et al., 2009; Woltjen et al., 2009). Transposons are mobile genetic elements occurring naturally in the genome, which essentially 'jump', or 'cut and paste' themselves, around the genome. The *piggyBac* transposable elements are not found in the mammalian genome, offering a system that could be utilized for introduction of reprogramming factors. The system harbours 2 main advantages over other transposable elements: transposition of plasmid into genome is highly efficient, and the transposon can subsequently excise itself with extraordinary precision allowing a 'foot-print free' transgene removal (Wang et al., 2008). Indeed, the Woltjen paper demonstrated not only the production of iPSCs, but near perfect excision of the transgene from the established iPSC lines (Woltjen et al., 2009).

1.2.1 Secondary reprogramming systems

While the polycistronic, lentiviral and *piggyBac* systems improved reprogramming efficiency, there still remained the challenge that each cell had different numbers of reprogramming cassettes inserted in various regions of the genome. Those systems produced a heterogeneous group of cells with different levels of reprogramming factor expression. To circumvent such challenges, groups began using established iPSC lines to generate so-called secondary reprogramming systems. Such systems rely on expression of the 4 factors under control of the doxycycline (dox) inducible tetracycline promoter (P_{TET}), thus allowing tight control over reprogramming factor expression. The dox system is based upon constitutive expression of the reverse tetracycline transactivator (rtTA) protein, which can only bind the tertracycline operator (*tetO*) and activate the P_{TET} promoter in the presence of doxycycline, producing a so-called Tet-ON system (Gossen and Bujard, 1992; Urlinger et al., 2000). In secondary reprogramming systems, the *tetO* and P_{TET} sequences are placed on the reprogramming cassette(s) upstream of the 4 factors, in conjunction with constitutive rtTA expression in the cells, allowing for dox dependent reprogramming factor induction. In one of the first secondary reprogramming systems developed, Maurius Wernig and colleagues took iPSC lines generated by dox inducible lentiviral plasmids, with rtTA expressed from the *Rosa26* locus, and generated chimeric mice from the clonal iPSC lines (Wernig et al., 2008). Upon isolation of various somatic tissue types from the chimeric mice, they demonstrated that the addition of doxycycline could activate the integrated reprogramming cassettes, initiating successful reprogramming. The secondary system exhibited 25-50-fold higher reprogramming efficiencies than the initially reprogrammed fibroblast lines (Wernig et al., 2008). Further studies have gone on to use transposon based secondary systems, with polycistronic reprogramming cassettes, including lines that have a single reprogramming cassette integration (O'Malley et al., 2013; Tonge et al., 2014; Woltjen et al., 2009, #468). Such efficient, highly reproducible, secondary reprogramming systems have enabled detailed analysis of the reprogramming process, from single cell to high throughput genome wide analyses, unlocking the basic

mechanisms of the reprogramming process. Yet, reprogramming efficiency is still underwhelming with less than 3% of the starting cells making it to a pluripotent state. Such low efficiency poses challenges when attempting to perform bulk population analyses, as only a small fraction of cells makes it to a pluripotent state. Thus, the time point and sub-population of cells used for any analyses must be carefully considered with regard to the question asked.

1.2.2 Cell surface markers to monitor reprogramming progression

Because the induction of pluripotency from differentiated cells is often asynchronous and inefficient, it is necessary to utilize cell surface markers and gene reporters to decipher sub-sets of cells en-route to pluripotency. Initially two core strategies were envisioned to investigate minor reprogramming populations amongst the mass of cells which fail to reprogram: either by monitoring the down-regulation of genes known to be expressed in lineage committed cells, and/or by following the gradual activation of genes associated with pluripotency. The challenge has been to identify what genes strictly represent the pluripotent state or the loss of a given differentiated state.

One of the first papers to attempt a combinatorial reporter strategy for identification of reprogramming intermediates monitored the loss of the fibroblast (and other differentiated cells) surface marker Thy1, gain of signal for the stem cell stage-specific embryonic antigen-1 (SSEA-1), in conjunction with expression of either *Oct4*- or *Sox2*-GFP reporters (Stadtfeld et al., 2008). When reprogramming from fibroblasts, the majority of cells down regulate Thy1 within 3 days of reprogramming, indicating that loss of *Thy1* is a common step, but not a major roadblock to reprogramming (O'Malley et al., 2013; Samavarchi-Tehrani et al., 2010). Gradual gain of SSEA1 expression (in Thy1- cells), and subsequent activation of the *Oct4/Sox2*-GFP reporter, allowed for isolation of populations that had increased probability to generate iPS colonies (Stadtfeld et al., 2008). However, it has subsequently been shown that SSEA-1 is expressed heterogeneously in ES cells, and its expression comes on quite early and heterogeneously during the reprogramming process, which precludes its use as a reliable

marker of early-stage reprogramming populations (Cui et al., 2004; O'Malley et al., 2013; Samavarchi-Tehrani et al., 2010; Stadtfeld et al., 2008). Regardless, these early works provided the framework for the hypothesis that reprogramming is a step-wise process, with cells gradually progressing to more advanced reprogramming states.

In 2010, two groups simultaneously published findings describing the acquisition of epithelial character as a defining feature of the early to mid-stages of reprogramming (Li et al., 2010a; Samavarchi-Tehrani et al., 2010). E-cadherin is critical in the establishment and maintenance of pluripotency, is uniformly expressed in ES cells, and can replace Oct4 as part of the 4 factor reprogramming strategy (Malaguti et al., 2013; O'Malley et al., 2013; Redmer et al., 2011). Markers of epithelial character, such as E-cadherin, can be used to identify early populations that are progressing to pluripotency, perhaps more stringently than SSEA-1 (Li et al., 2010a; Samavarchi-Tehrani et al., 2010).

Until recently, there remained a rather large gap in our ability to distinguish the mid-to-final stages of reprogramming before cells fully engage the core pluripotency network. A recent study performed time-course microarray analysis of a *piggyBac* transposon based secondary reprogramming system and revealed the expression of cell surface markers CD44 and ICAM1 (CD54) markedly change during the reprogramming process (O'Malley et al., 2013). During reprogramming, ICAM1 begins heterogeneously expressed in MEF but then quickly becomes low by day 2, and is then up-regulated at the late stages of reprogramming concomitantly with many pluripotency genes. Whereas CD44 expression begins uniformly high in fibroblasts and is gradually down-regulated over the initial-to-mid stages of reprogramming, closely following activation of E-cadherin expression. The two surface markers, CD44 and ICAM1, paired with an endogenous *Nanog*-GFP reporter, were utilized to isolate various sub-populations of the reprogramming process. As cells lose CD44 expression, and gain ICAM1 expression, they correspondingly have greater propensity to generate iPSC colonies, as assessed by sorting and re-plating reprogramming experiments (O'Malley et al., 2013). Single cell mass cytometry analysis also demonstrated ICAM1 as a robust surface marker of successful reprogramming (Lujan et al., 2015; Zunder et al., 2015). Furthermore, global

RNA-seq expression profiles of the successive ICAM1/CD44 populations identified intermediate and late populations of cells that progressively gain global transcriptional profiles similar to ES/iPSCs (O'Malley et al., 2013). Such analyses have elucidated reprogramming intermediates and highlighted the distinct stages of reprogramming. It is now accepted that conventional 4 factor reprogramming is a step-wise process, involving intermediate populations along the path to pluripotency.

1.2.3 Gene reporters to assess acquisition of pluripotency

Often reprogramming efficiency and dynamics are monitored and interpreted based upon pluripotency gene reporter expression. However, researchers use caution when relying upon a single marker of the pluripotent state given the fact that most, if not all, pluripotency genes are not limited to expression in stem cells. A classic example would be Sox2; a gene that is a crucial part of the core pluripotency network in a triad with Oct4 and Nanog, but also plays a major role in maintaining and specifying neural lineages (Gagliardi et al., 2013; Rizzino, 2013; Thiel, 2013; Wang et al., 2012). Recent work has also demonstrated lineage-directing roles for Oct4 and Nanog (Karwacki-Neisius et al., 2013; Wang et al., 2012). Because there is no consensus on a gene restricted strictly to expression in the pluripotent state, to be certain the resultant cells are indeed iPSCs, it is often necessary to perform other tests such as extended passages, transcriptional profiling and various differentiation assays to assess ability to generate the three germ layers.

The most commonly used reprogramming reporter is Oct4, a master regulator for the induction and maintenance of pluripotency (Jerabek et al., 2014; Scholer et al., 1989; Takahashi et al., 2007; Takahashi and Yamanaka, 2006). While Oct4 directly binds and regulates most of the core regulatory network in ES cells (Chen et al., 2008; Dunn et al., 2014; Mullen et al., 2011), it is expressed heterogeneously and at high levels during differentiation to multiple cell lineages (Karwacki-Neisius et al., 2013; Niwa et al., 2000; Wang et al., 2012) (Radzisheuskaya et al., 2013). A close look at time-course expressional analysis during reprogramming, including in-depth single-cell analysis, has

revealed that endogenous Oct4 is expressed early in the process, before Sox2 or Nanog, and at or near ES cell levels in intermediate populations and partially reprogrammed iPSC colonies (Buganim et al., 2012; Greder et al., 2012; O'Malley et al., 2013; Polo et al., 2012). Therefore, using Oct4 as a marker of the pluripotent state may over-estimate reprogramming efficiency and mistake partially reprogrammed populations as pluripotent.

Despite its role in neural stem cells and neurogenesis, Sox2 represents a good reporter of pluripotency during reprogramming. Sox2 gene expression is restricted to the end stages of reprogramming, with few cells demonstrating expression at intermediate stages of reprogramming (Buganim et al., 2012; O'Malley et al., 2013). The activation of endogenous Sox2 expression marks a distinct phase of reprogramming marking initial establishment of the pluripotency network and maturation toward stable iPSCs (Buganim et al., 2012). Sox2 might operate upstream of well-established late stage pluripotency associated genes such as Esrrb, Sall4 and Lin28, as suggested by Bayesian modelling and over-expression of the downstream factors allowing for reprogramming in the absence of Sox2 (Buganim et al., 2012). Recent work using Sox2-EGFP reporter MEF demonstrated the protein as a reliable reporter of the pluripotent state and much more stringent than an Oct4-EGFP reporter (Lujan et al., 2015).

Perhaps the most reliable reporter of the fully reprogrammed state is Nanog, a core pluripotency network protein. It is expressed late during the reprogramming process (Buganim et al., 2012; Greder et al., 2012; O'Malley et al., 2013), is important for establishment of pluripotency in the early embryo (Silva et al., 2009) and can drive completion of reprogramming in minimal conditions or transition to naïve pluripotency in partially reprogrammed (pre-iPS) cells (Moon et al., 2013; Silva et al., 2009; Theunissen et al., 2011). However, investigations of the various stages of reprogramming identified detectable Nanog expression in a subset of early reprogramming populations (Buganim et al., 2012; O'Malley et al., 2013). Furthermore, Nanog is not required for the maintenance of pluripotency, and recent work has demonstrated that Nanog null cells can reprogram to iPSCs, albeit with lower efficiency (Carter et al., 2014; Schwarz et al., 2014). While Nanog may not be absolutely necessary

for pluripotency acquisition (in vitro) nor maintenance, it may represent the best-tested marker of late stage reprogramming and maturation to the pluripotent state.

Recent work has highlighted candidate genes that might serve as reliable reporters of different reprogramming stages. Time course mass cytometry analysis, probing cell surface marker expression during reprogramming, identified CD77 and CD49 as transiently expressed genes that mark early reprogramming populations with increased probability to successfully reprogram. Those novel surface markers also enabled the identification of gene expression profiles of reprogramming prone populations, revealing the previously identified pluripotency genes Nrob1 and Etv5 as early up-regulated genes whose expression is important for reprogramming progression (Lujan et al., 2015). Such genes represent good candidates to explore early reprogramming populations in search of the early events required prior to establishment of the complete pluripotency network. Single cell analysis of 7 distinct stages of reprogramming, including 9 clonal populations of partially reprogrammed cell-lines and 5 established iPS lines, has identified genes which are highly expressed only in late stage reprogramming and mature iPSCs. The work demonstrated factors such as Dppa2, Esrrb, Lin28 and Utf1 to be uniformly activated in late stages of reprogramming and in colonies representing bona fide iPSCs (Buganim et al., 2012). However, as highlighted by single-cell mass cytometry analysis, Lin28 may also mark an alternative reprogrammed state, that has yet to be well characterized (Zunder et al., 2015). Regardless, the aforementioned genes, or a combination of reporters, may represent the best strategy if one seeks to determine the key players involved in establishing a stable, uniform network of pluripotent cells.

1.3 The hallmark stages of reprogramming from fibroblasts

1.3.1 Senescence

Years after the discovery of iPSCs, despite reprogramming technologies improving, the best reported efficiencies remained around 1% of starting cells, highlighting the

existence of major unidentified roadblocks. In the summer of 2009, 4 groups simultaneously published papers describing senescence as a major reprogramming roadblock (Banito et al., 2009; Hong et al., 2009; Kawamura et al., 2009; Li et al., 2009). Senescence is defined as the irreversible arrest of cell growth, and is often associated with the regulation of the *INK4a/Arf* locus, which encodes 2 critical senescence-inducing proteins (Collado et al., 2007). The two proteins are expressed at moderate levels in MEF and almost undetectable in ES or iPS cells (Li et al., 2009). Within 24 hours of reprogramming factor induction, up-regulation of the core senescence network genes $p16^{INK4a}$, *p53*, and $p21^{CIP1}$ occurs (Banito et al., 2009). There does not appear to be just one reprogramming factor responsible for senescence induction, as independent expression of Sox2, Klf4 or cMyc alone is capable of activating expression of the senescence pathway (Banito et al., 2009). As reprogramming progresses, the levels of senescence network proteins transition to lower levels of detection (Banito et al., 2009; Li et al., 2009). Those findings suggest that only cells that avoid senescence gene up-regulation, or ones that successfully down-regulate them can reprogram. Indeed, knockdown of $p16^{Ink4a}$, $p19^{ARF}$, *p21* or *p53*, or reprogramming of the corresponding knockout MEFs, results in up to 15-fold increase in reprogramming efficiency (Banito et al., 2009; Kawamura et al., 2009; Li et al., 2009). Strikingly, knockdown or knockout of *p53* negates the requirement for *Klf4* and *c-Myc* in the reprogramming cocktail, enabling successful reprogramming with just *Oct4* and *Sox2* (Hong et al., 2009; Kawamura et al., 2009). Presently, senescence remains one of the best-characterized roadblocks to reprogramming, and is commonly targeted for down-regulation in robust reprogramming protocols.

1.3.2 Sequential activation of pluripotency genes

Very early in the field of reprogramming research, as one might predict, it became apparent that the transition to a pluripotent state happens sequentially in a step-wise manner. Work from Rudolf Jaenisch's lab provided the first evidence, showing the pluripotency-associated marker alkaline phosphatase (AP) is activated early and before

stage-specific embryonic antigen 1 (SSEA-1). Both AP and SSEA-1 mark the early events in the reprogramming process, and precede activation of the pluripotency genes *Oct4* and *Nanog* (Brambrink et al., 2008; Stadtfeld et al., 2008). More comprehensive, time-course analyses of the sequential nature of reprogramming surfaced years later; describing 2 primary waves of pluripotency gene activation, including a late maturation phase (Buganim et al., 2012; O'Malley et al., 2013; Polo et al., 2012). While the exact timing differs between systems, it has been observed that pluripotency associated genes such as *Oct4, Sall4, Sall1, Fgf4* are expressed during the first wave at the mid-stages of reprogramming, after MET, while *Nanog, Sox2, Dppa* family members, *Esrrb, Lin28* and *Klf2* are expressed in a second wave late in the process (Buganim et al., 2012; Golipour et al., 2012; O'Malley et al., 2013; Polo et al., 2012). Many other genes, which are not as well characterized in stem cells, fall into one of the two classes of intermediate and late stage activation during reprogramming, and require further scrutiny as to the relevance of their expression.

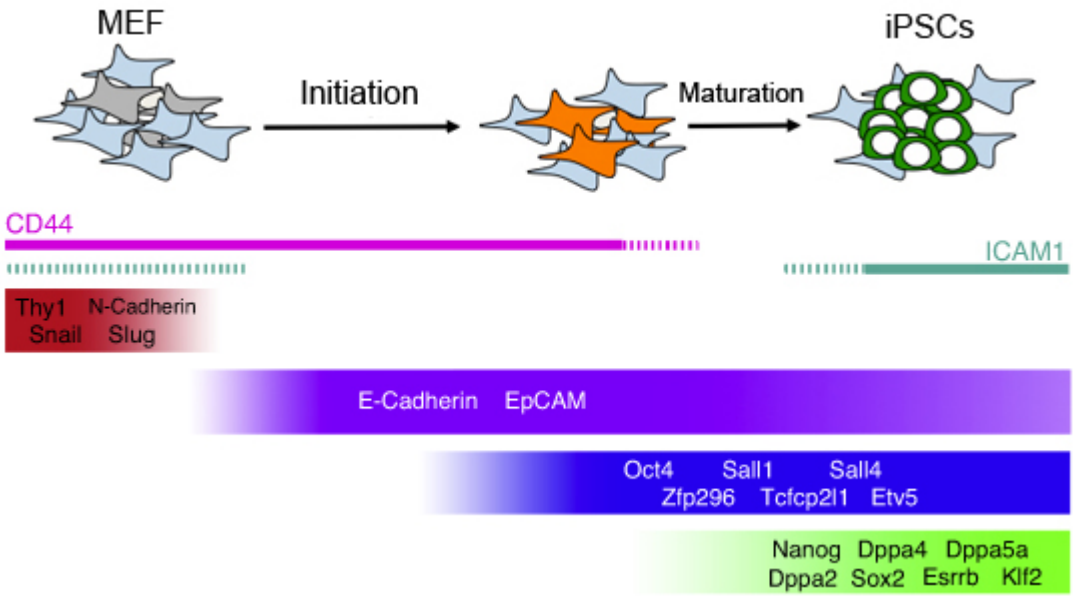

Figure 1.1 Stages of reprogramming

Early after initiation of reprogramming, fibroblast associated genes such as Thy1, N-cadherin, Snail and Slug are down-regulated, after which cells gain expression of the epithelial markers E-cadherin and EpCAM. Activation of the pluripotency network happens later in two general waves: the first containing hallmark genes such as Oct4 and Sall4, while the late maturation phase is characterized by expression of the core pluripotency network genes Nanog and Sox2 among others.

The final stages of reprogramming involve the silencing of the exogenous reprogramming factors, as cells transition to a more stable pluripotent state (Golipour et al., 2012; Maherali et al., 2007; Okita et al., 2008; Papapetrou et al., 2009; Wernig et al., 2007). It was initially observed that high levels of reprogramming factor expression lock cells in a unique state, whereby they do not give rise to the three germ layers in teratomas (Brambrink et al., 2008). To concisely investigate the so-called 'maturation phase' of reprogramming, Galipour and colleagues utilized their dox inducible reprogramming system to study the affects of dox withdrawal at various end points of reprogramming (Golipour et al., 2012). They found that maintenance of high levels of MKOS, up to 25 days post induction, could impede activation of late pluripotency genes such as *Dppa* family members, *Lin28, Sox2* and *Utf1*. Within 3 days of dox withdrawal and 4-factor down-regulation, the late pluripotency genes were robustly up regulated. Indeed, an siRNA screen at day 20 of reprogramming, targeting the genes which were up-regulated upon dox withdrawal, resulted in up to 50% reduction in $DPPA4^+$ colonies, highlighting the importance of the maturation phase (Golipour et al., 2012). In further support, the Yamanaka lab recently reported that, while 20% of transfected human fibroblasts express the pluripotency surface marker *Tra-1-60* within 7 days of reprogramming initiation, only 1% of those become stable, expandable iPSC lines (Tanabe et al., 2013). Over-expression of the late pluripotency gene *Lin28*, enhanced stability of sorted and reseeded $TRA1\text{-}60^+$ cells, confirming the involvement of such late pluripotency genes in iPSC stability (Tanabe et al., 2013). A subsequent paper performing time course analysis of human fibroblast reprogramming also confirmed distinct activation of late pluripotency associated genes such as *Lin28, Sox2* and *Dnmt3b*, which inversely correlated with the silencing of the reprogramming factor

transgenes (Takahashi et al., 2014). Recent work has also identified another unique cell population of partially reprogrammed cells associated with high levels of the reprogramming factors, termed F-class stem cells, for there 'fuzzy' appearance under the microscope (Tonge et al., 2014). F-class cells are morphologically and transcriptionally distinct from ESCs, but exhibit pluripotent characteristics such as the ability to produce teratomas. While the F-class cells are by definition pluripotent, they are reliant upon continued transgene expression, and are unable to contribute to chimeras and fail to incorporate into blastocysts upon injection. Thus, the maturation phase and generation of pluripotent cells that can give rise to chimeric animals, is reliant upon transgene silencing/down-regulation, representing a considerable roadblock and important step toward stable pluripotency.

1.3.3 Mesenchymal to epithelial transition

As reprogramming progresses, one of the earliest observed changes is the loss of fibroblast-associated genes and gain of epithelial genes expression, termed a mesenchymal-to-epithelial transition (MET). The fibroblast marker *Thy-1* silencing happens rapidly and near completely very early in the reprogramming process (Li et al., 2010a; O'Malley et al., 2013; Polo et al., 2012; Samavarchi-Tehrani et al., 2010; Stadtfeld et al., 2008). In addition to *Thy-1*, many other mesenchyme-associated genes are robustly silenced early in the reprogramming process, including *Snai1, Slug, Cdh2, Zeb1, Zeb2* among others (Li et al., 2010a; O'Malley et al., 2013; Samavarchi-Tehrani et al., 2010). While mesenchymal associated genes are silenced, epithelial genes such as *E-cadherin* (*E-cad*), multiple claudins, epithelial cell adhesion molecule (Ep-CAM) and *occludin* are robustly activated (Li et al., 2010a; Samavarchi-Tehrani et al., 2010). Correspondingly, knockdown or over-expressing E-cad, addition of BMP ligands, or over-expression of the micro RNA (miRNA) clusters 200, 205 or 302-367, improves reprogramming efficiency through mechanisms hypothesized to enhance the MET (Chen et al., 2010; Ichida et al., 2009a; Li et al., 2010a; Liao et al., 2011; Samavarchi-Tehrani et al., 2010). However, in each study it is not clear if the observed increase in MET is

cause or consequence of increased reprogramming efficiency. While MET has proven a hallmark of the early stages of reprogramming, the transition happens quite efficiently in robust reprogramming systems (O'Malley et al., 2013), and therefore might not be considered a major roadblock to reprogramming n all systems. However, inhibition of TGF-β receptors, a pro-mesenchymal signalling pathway, remarkably enhances reprogramming efficiency, suggesting the MET is a barrier to successful reprograming. Yet inexplicably, TGF-β inhibitors also improve reprogramming efficiency of epithelial cells and promote reprogramming when added at late stages of the process, long after the MET has occurred (Ichida et al., 2009a; Li et al., 2010a). It is therefore evident that the contribution of TGF-β signalling during reprogramming requires further investigation.

1.4 TGF-β signalling in stem cells and reprogramming

One of the single most powerful small molecules that can be added to the reprogramming media is a TGF-β type I receptor inhibitor, which results in up to 60-fold increase in reprogramming efficiency (Ichida et al., 2009b; Li et al., 2010a; Maherali and Hochedlinger, 2009). TGF-β signalling is involved in many cellular and developmental processes, ranging from maintenance of pluripotency, formation of the primitive streak, epithelial to mesenchymal transitions (EMT), lineage specification, apoptosis, senescence and proliferation (Derynck and Zhang, 2003; Oshimori and Fuchs, 2012; Shi and Massague, 2003a). Because of the well established roles for TGF-β in EMT, it has been proposed that TGF-β receptor inhibition blocks pro-mesenchymal signalling during reprogramming, allowing cells to progress toward the epithelial-like pluripotent state (Polo and Hochedlinger, 2010). While TGF-β signalling is involved in mesenchymal signalling, it effects multiple other cellular states, which may also contribute to the inhibitors affect on the reprogramming process. The following sections aim to illuminate the various cellular functions of TGF-β signalling that might affect reprogramming.

1.4.1 Overview of TGF-β signalling

Initiation of TGF-β signalling is accomplished by a superfamily of ligands consisting of over 30 members, each of which binds specific sub-classes of TGF-β associated cell surface receptors (See Figure 4.1 on page 77). Signalling commences upon ligand receptor interaction, initiating type II receptors to associate with, phosphorylate and thereby activate Type I receptors. Upon activation, Type I receptors initiate downstream signalling with their serine/threonine kinase domains by phosphorylating cytoplasmic cofactors such as Smad2/3, MAPK, PI3K/Akt, RhoA, TAK1/MEKK1 (Derynck and Zhang, 2003; Oshimori and Fuchs, 2012; Shi and Massague, 2003a; Suwanabol et al., 2012). Smad transcription factors are perhaps the most instrumental in executing TGF-β receptor signalling. Loss of Smad protein in cells severely diminishes or completely abolishes the TGF-β induced response (Li et al., 2013; Piek et al., 2001; Sakaki-Yumoto et al., 2013). Smads are expressed throughout almost all cells and tissues of the body, with signalling regulation occurring via post-transcriptional modification (Derynck and Zhang, 2003; Oshimori and Fuchs, 2012; Shi and Massague, 2003a). Receptor Smad proteins (R-Smad's) are phosphorylated and activated by TGF-β receptors. Upon phosphorylation, R-Smad's form hetero or homodimers and complex with the co-activator Smad4, which stabilizes the interaction and facilitates co-factor associations and DNA binding within the nucleus (Chen et al., 1997; Xi et al., 2011). Within just 1-2 hours of TGF-β stimulation, Smads have transported to the nucleus and elicit differential expression of hundreds of genes (Kang et al., 2003; Levy and Hill, 2006; Mullen et al., 2011).

1.4.2 Smad mediated signalling

There are 2 main branches of Smad signalling: SMAD2/3 and SMAD1/5/8. SMAD2/3 signalling is initiated by TGF-β ligand subfamilies such as TGF-β's, activin and nodal, whereas SMAD1/5/8 signalling is initiated by the ligand subfamilies BMPs (bone morphogenetic proteins), GDFs (growth and differentiation factors) and MIS (Muellerian inhibiting substance)(Shi and Massague, 2003b). All Smads contain a

highly conserved c-terminus MH2 domain with a consensus serine-X-serine (SXS) motif, which is the site of phosphorylation by ligand-activated receptors (Abdollah et al., 1997; Souchelnytskyi et al., 1997). The C-terminal phosphorylated active Smads form complexes with other R-Smads and Smad4 and then translocate to the nucleus. Once in the nucleus, Smads act as bipartite transcription factors, working as part of transcription regulatory complexes.

SMAD signalling inactivation is accomplished largely through the inhibitor Smads (I-Smads), SMAD6 and SMAD7. Through direct binding to TGF-β type I receptors, SMAD6 and SMAD7 block Smad-activating phosphorylation and also facilitate type I receptor ubiquitination and degradation (Derynck and Zhang, 2003; Ebisawa et al., 2001; Kavsak et al., 2000). Smad6 can also directly interfere with SMAD1/5/8 association with SMAD4, thereby destabilizing the BMP signalling pathway. Additionally, SMAD7 can bind DNA at Smad response elements and interfere with SMAD2/3/4 interactions with the genome (Zhang et al., 2007). Both SMAD6 and SMAD7 are activated transcriptionally by SMAD2/3 or SMAD1/5/8, producing auto-regulatory negative feedback loops. Furthermore, TGF-β1 ligand stimulation in cells initiates rapid SMAD7 translocation to the plasma membrane in complex with the E3 ubiquitin ligase Smurf, which then target TGF-β type I and type II receptors for degradation (Itoh et al., 1998; Kavsak et al., 2000). The regulation of Smad6/7 allows for antagonistic competition between the Smad2/3 and Smad1/5/8 signalling pathways. For example, inhibition of SMAD2/3 phosphorylation, results in an increase in SMAD1/5 phosporylation through a reduction of Smad7 activity (Galvin et al., 2010). Conversely, addition of TGF-β ligand to cell cultures down-regulates expression of BMP signalling through repression of BMP ligands and receptors (Gronroos et al., 2012; Kang et al., 2003; Xu et al., 2008). Thus, there is reciprocal competition amongst the Smad2/3 and Smad1/5/8 signalling pathways producing an antagonistic relationship.

1.4.3 Smad2/3 mediated transcriptional activation

The binding affinity of Smads to DNA is weak, roughly ~100 fold less than other high affinity transcription factors (Shi et al., 1998), and therefore Smad-cofactor interactions are crucial in specifying interactions with the genome (Derynck and Zhang, 2003). The consensus Smad binding element (SBE), 5'-AGAC-3', is common throughout the genome and often found in close proximity to recognition sequences of other transcription factors (ten Dijke et al., 2000). Despite Smads having low binding affinity for the SBE, somewhat inexplicably, their DNA binding is still required for transcriptional regulation (Jones and Kern, 2000; Takaesu et al., 2005). In such a model, transcription factors recruit Smads to their target loci, Smads then binds the SBE while in a complex with the transcription factor and cofactors, which can then remodel chromatin and/or regulate transcription (Derynck and Zhang, 2003). The reliance on cofactors for recruitment and binding of Smads to DNA produces a transcriptional regulatory network that is extremely context dependent (Dennler et al., 1999; Mu et al., 2012; Shi et al., 1998; Zawel et al., 1998). The widely conserved Smad C-terminal MH2 domain enables interaction with several transcriptional regulators and nucleosome remodelers such as p300/CBP, Mediator, MLL, SWI/SNF, TRIM33 and the demethylase JMJD3 (Bertero et al., 2015; Estaras et al., 2012; Kato et al., 2002; Xi et al., 2008; Xi et al., 2011). Perhaps the best-characterized Smad3-cofactor interaction is with the p300 transcriptional activator (Pouponnot et al., 1998). Smad3 couples with p300 as part of the enhanceosome complex, which drives recruitment of cofactors and RNAPolII to active sites of transcription (Vo and Goodman, 2001). P300 also demonstrates histone acetyltransferase (HATs) activity, which helps remodel nucleosomes at target gene promoters to activate transcription (Sterner and Berger, 2000). The TGF-β transcriptional response is largely dependent upon SMAD3 interaction with p300 (Pouponnot et al., 1998). Over-expression of *E1A*, which competes with SMAD3 for p300 binding, inhibits Smad3 dependent transcriptional response to TGF-β stimulation. Furthermore, over-expression of p300, or the Smad3-

p300 cofactor FOXH1, can boost SMAD3 transcriptional activation (Pouponnot et al., 1998).

Smad cofactors often orchestrate transcriptional activation through remodelling chromatin at target loci (Gaarenstroom and Hill, 2014). Indeed, most of the well-established Smad cofactors such as p300, mediator, SWI/SNF and Jmjd3 exhibit chromatin remodelling ability. An example of Smad3 assisted chromatin remodelling comes from investigations of its interaction with the demethylase JMJD3. SMAD3 and JMJD3 cooperate in chromatin remodelling in a number of different cellular contexts including regulation of pluripotency, and during endoderm and neural developmental programs. As with most cell types, the SMAD3 binding profile in hESCs and endoderm are almost completely unique. As cells exit pluripotency and enter the endoderm differentiation program, SMAD2/3 with JMJD3 are localized to endoderm specific genes, JMJD3 removes the repressive histone H3 Lys27 (K27) trimethylation (H3K27me3) marks, poising endoderm genes for activation (Dahle et al., 2010). The recruitment of JMJD3 to endoderm genes requires active SMAD2/3 signalling, and inhibition of TGF-β receptors blocks JMJD3 recruitment (Dahle et al., 2010). Similarly, in neural stem cell, SMAD3 recruits JMJD3 to neural lineage specific promoters, coordinating the neural developmental program both *in vitro* and in the developing chick embryo (Estaras et al., 2012). Within 30 minutes of TGF-β stimulation in neural stem cells, JMJD3 is recruited to neural gene promoters, and this response is completely lost in the context of *Smad3* knock down. The SMAD3-JMJD3 bound genes exhibit H3K27me3 reduction within 3 hours of TGF-β activation, and a corresponding increase in mRNA expression levels (Estaras et al., 2012). An eloquent *in-vivo* experiment demonstrated that electroporation of a plasmid encoding expression of a constitutive active Smad3 (Smad3CA) in the neural tube of chick embryos resulted in enhanced neurogenesis from resident neural stem cells. This Smad3CA phenotype was reliant upon Jmjd3, as indicated by loss of the phenoypte in the context of *Jmjd3* knockdown (Estaras et al., 2012). Such experiments have conclusively demonstrated roles for Smad3 and Jmjd3 cooperating in remodelling chromatin to activate transcriptional programs in

diverse cellular contexts, assisting in both the endodermal and ectodermal specification programs.

1.4.4 Developmental roles for TGF-β signalling

TGF-β signalling first came to the spotlight in the developmental biology field when a retroviral insertional mutagenesis screen identified null mutants of *Nodal*, a TGF-β family member ligand, to have an embryonic lethal phenotype (Conlon et al., 1991). Nodal-null embryos arrest at the gastrulation stage of development, showing no signs of formation of a primitive streak (Conlon et al., 1994; Zhou et al., 1993). Wild-type ES cells, at very low contribution, can rescue gastrulation in nodal null chimeric embryos, suggesting that minimal, localized Nodal signalling is sufficient to initiate a primitive streak and gastrulation (Varlet et al., 1997b). Nodal transcript is initially detected at E5.5 in the primitive ectoderm, then localizes to the posterior region of the epiblast, and also within the visceral endoderm (Conlon et al., 1994; Varlet et al., 1997b). High contribution ES cell chimeric rescue (>80% WT cells embryo) of nodal-null embryos, revealed that deficient Nodal signalling in the primitive endoderm results in anterior axis defects, with embryos exhibiting almost complete absence of the forebrain among other anterior neural defects (Varlet et al., 1997b). Because Nodal is required for the formation of the primitive streak, it became an attractive hypothesis that Nodal/TGF-β signalling was a master regulator of mesoderm fate specification. However, it is clear that Nodal null cells do give rise to many mesoderm populations in chimeric embryos (Conlon et al., 1991; Conlon et al., 1994; Varlet et al., 1997a; Varlet et al., 1997b). Furthermore, Nodal null embryos exhibit marked defects in cells and tissue of all three germ layers, suggesting Nodal/TGF-β signalling is involved in many unique developmental contexts that are not limited to mesodermal signalling (Conlon et al., 1994; Varlet et al., 1997b). With a strikingly similar phenotype, deletion of the Nodal downstream target Smad2, results in embryonic lethality due to a gastrulation defect (Nomura and Li, 1998). *Smad2* null embryos fail to form an organized egg cylinder and falter in formation of the mesoderm (Nomura and Li, 1998). Conditional *Smad2* null

embryos, where the deletion is epiblast specific, can form a primitive streak and undergo normal gastrulation, with defects arising in anterior neural patterning in a similar phenotype as nodal mutants (Vincent et al., 2003). Thus, like Nodal embryonic roles, Smad2 function can be attributed to the extraembryonic signalling centre within the anterior visceral endoderm (Heyer et al., 1999; Tremblay et al., 2000; Waldrip et al., 1998). Intriguingly, null mutation of the highly homologous Smad3, sharing 92% amino acid identity with Smad2, does not result in embryonic lethality. *Smad3* null embryos give rise to juvenile mice, albeit with severe immune dysfunctions, endoderm lineage defects and metastatic colorectal cancer in 100% of mice examined (Weinstein et al., 2001; Yang et al., 1999; Zhu et al., 1998). The strikingly different phenotypes of *Smad2* and *Smad3* knockouts, despite high sequence identity and shared regulatory pathways, appear to be partly a result of differential expression patterns in the embryo. *Smad2*, but not *Smad3*, is expressed in the anterior visceral endoderm, where Smad2 is essential in providing the signalling to initiate gastrulation and coordinate anterior-posterior patterning (Tremblay et al., 2000). In a single eloquent experiment, Elizabeth Robertson's group demonstrated that inserting *Smad3* in the *Smad2* null locus, rescues the *Smad2* knockout gastrulation defect, producing viable, fertile mice, albeit with reduced survival (10% *Smad3* rescued *Smad2* null embryos, as compared to the expected 25% Mendelian ratio)(Dunn et al., 2004). Those data thereby support the hypothesis that regional *Smad2/Smad3* embryonic expression facilitates their unique developmental roles, and further that SMAD2 and SMAD3 are at least partially functionally redundant.

1.4.5 TGF-β signalling in stem cells

Clues as to why TGF-β signalling modulation affects reprogramming may come from studies examining the pathways' role in stem cells. Most studies of Smad2/3 in stem cells come from work with human ES cells (hESCs) or the developmentally comparable mouse epiblast stem cells (mEpiSCs). Mouse EpiSCs are different from conventional mESCs in that they are derived from the post implantation epiblast, rather

than the pre-implantation embryo, representing a more primed pluripotent state (Tesar et al., 2007). Both mEpiSCs and hESCs are propagated in culture independent of LIF, and rely on Fgf and TGF-β signalling for the maintenance of pluripotency (Beattie et al., 2005; Brons et al., 2007; James et al., 2005; Tesar et al., 2007; Vallier et al., 2005). In hESCs TGF-β/activin signalling initiates SMAD2/3 localization to the *Nanog* promoter to drive transcription (Xu et al., 2008). Mutation of the putative *Smad2/3* binding elements (SBEs) in the *Nanog* proximal promoter reduces promoter activity and responsiveness to activin induced *Nanog* expression (Xu et al., 2008). Furthermore, depletion of Smad2 and/or addition of TGF-β inhibitors, results in abrupt exit from a pluripotent state and activation of lineage specific genes (Beattie et al., 2005; James et al., 2005; Sakaki-Yumoto et al., 2013; Vallier et al., 2005).

In the more naïve mESCs, chromatin immunoprecipitation (ChIP) experiments revealed that SMAD3 is bound to >80% of loci bound by the core pluripotency transcription factors OCT4, SOX2 and NANOG. Furthermore, a brief 1-hour activation or repression of TGF-β signalling, results in significant up or down regulation of the SMAD3 bound genes, indicating transcriptional regulatory roles for SMAD3 in mESCs (Mullen et al., 2011). However, evidence points to a negative regulation of self-renewal by SMAD2/3 in naïve mES cells, as addition of TGF-β ligands induce differentiation and inhibitors of TGF-β promote more homogenous *Nanog* expression and self-renewal (Galvin-Burgess et al., 2013; Xi et al., 2011; Ying et al., 2003). It remains unclear why the response to TGF-β signalling is completely contradictory when comparing mouse naïve and EpiSCs.

The TGF-β antagonistic BMP pathway promotes differentiation in hESCs and mEpiSCs (Pera et al., 2004; Xu et al., 2002; Xu et al., 2005). Treatment of hESCs with BMP2, BMP4 or BMP7 drives the cells to differentiate toward various lineages including endoderm (BMP-2, -4, or -7 addition), or trophoblast-like cells (BMP-4) (Mummery et al., 2003; Xu et al., 2002). Treatment with the BMP inhibitor NOGGIN allows for expansion and maintenance of an undifferentiated state in the absence of fibroblast feeders, which at the time were required to hESC maintenance (Xu et al., 2005). The BMP downstream targets SMAD1/5/8 bind the *Nanog* promoter in hESCs

within 24 hours of BMP4 induction, coinciding with exit from an undifferentiated state, suggesting they may act in an inhibitory role at the *Nanog* promoter (Xu et al., 2008).

The roles of TGF-β and BMP signalling in self-renewal are almost completely, and somewhat inexplicably reversed when investigating naïve mES cells. Foremost, mouse ES cells are maintained in the absence of feeders with LIF and BMP4 (Galvin-Burgess et al., 2013; Ying et al., 2003). Ying et al first identified that BMP4 maintains pluripotency by blocking ES cell differentiation into neural lineages (Ying and Smith, 2003). They subsequently demonstrated that BMP4 (or GDF6, another SMAD1/5/8 activating ligand) and Leukemia inhibiting factor (LIF) alone were sufficient to maintain pluripotency in the absence of serum. The mechanism by which BMP acts is in part through up regulation of Id proteins, which bind to and block the activity of some of the bHLH family of pro-neural transcription factors (Ying et al., 2003). Other groups have also demonstrated that addition of BMP4 suppresses Mitogen activated protein kinase (MAPK) and extracellular receptor kinase (ERK) pathways, which also helps prevent differentiation (Qi et al., 2004a). Thus, it appears that rather than maintaining the pluripotency network of ES cells, BMP's play important roles in dampening differentiation cues. Interestingly, over-expression of *Nanog* superseded the requirement for BMP signalling to maintain ES pluripotency in serum-free media. Furthermore, *Nanog* over-expression is sufficient to activate Id genes (Ying et al., 2003). These findings bring to question the targets of BMP signalling in ES cells, specifically questioning whether BMP signalling directly induces Id expression, synergizes with NANOG to activate Id's, or indirectly activates the upstream pluripotency network.

1.4.6 Modulating TGF-β signalling during reprogramming

In the reprogramming context, modulation of TGF-β (SMAD2/3) and BMP (SMAD1/5/8) signalling has resulted in conflicting reports. Initial reports demonstrated that addition of BMP during reprogramming results in a 3-fold increase in reprogramming efficiency (Samavarchi-Tehrani et al., 2010). They further demonstrated

that BMPs act in part by activating the microRNA (mRNA) 200 family members, which are involved in promoting the mesenchymal to epithelial transition (MET) (Samavarchi-Tehrani et al., 2010). However, those results relied upon the early expressed reprogramming marker alkaline phosphatase staining as a read-out of successful reprogramming. More comprehensive analyses revealed that BMPs, present in the fetal calf serum (FCS) of reprogramming media, lock cells in a partially reprogrammed state and that inhibition of BMP signalling allows the so called pre-iPS cells to convert to mature bona fide iPSCs (Chen et al., 2013b). BMPs activate methyltransferases, such as Setdb1 and Suv39h1, which maintain repressive histone 3 lysine 9 methylation marks at pluripotency loci, acting as a barrier to reprogramming (Chen et al., 2013b). While BMPs may beneficially promote the early stage MET, at later stages of reprogramming the BMP pathway must be silenced to promote and stabilize the pluripotent state.

Investigations of the SMAD2/3 pathway have demonstrated that, in general, activation with TGF-β ligands reduces reprogramming efficiency while use of small molecules to inhibit the pathway results in a remarkable increase in reprogramming efficiency (Ichida et al., 2009a; Li et al., 2010a; Maherali and Hochedlinger, 2009). The Hochedlinger lab demonstrated that the TGF-β type I receptor inhibitor Alki (Alk5i), could increase reprogramming efficiency up to 60-fold (Maherali and Hochedlinger, 2009). The Alki was most effective when added during the first 8 days of reprogramming, exhibiting little effect when added at late time-points (Maherali and Hochedlinger, 2009). However, Ichida and colleagues demonstrated that, when reprogramming with just Oct4, cMyc and Klf4, the optimal time to add the Alki was day 10-22, showing improved efficiency over inhibition at early stages or during the entire reprogramming process (Ichida et al., 2009a). Therefore, not surprisingly, it seems TGF-β signalling during reprogramming is context dependent. The inhibitor was proposed to alleviate mesenchymal signalling, promoting MET at the early stages of reprogramming (Polo and Hochedlinger, 2010). However, addition of the Alki to epithelial cells still exhibits a 2-10 fold increase in reprogramming efficiency, suggesting other mechanisms are involved (Li et al., 2010a). Interestingly, alternative approaches to inhibit the TGF-β pathway with receptor blocking antibodies, does not

have near the positive affect as the Alki or SB43 receptor inhibitors (Ichida et al., 2009a). Furthermore, over-expression of TGF-β receptors, to boost pathway activation, only mildly inhibits the reprogramming process (Li et al., 2010a). Complicating matters further, is the recent evidence that exposure to TGF-β ligand for the first 3 days of reprogramming increases the generation of IPSCs by 2 to 3 fold (Liu et al., 2013). It seems, in a similar fashion to BMPs, TGF-β signalling may provide some benefit in early-intermediate reprogramming populations but must be silenced to reach the final pluripotent state. In any case, it is evident that modulation of TGF-β/BMP and Smad signalling has substantial implications to the outcome of reprogramming experiments and deserves further investigation.

1.4.7 Master transcription factors direct Smad3 to target loci

TGF-β stimulation or repression results in over 2-fold differential expression of several hundred genes within hours of treatment (Kang et al., 2003; Levy and Hill, 2006; Mullen et al., 2011). The genes and pathways targeted are highly context dependent, ranging from cell proliferation, senescence and apoptosis to maintenance of pluripotency and activation of development transcriptional programs (Derynck and Zhang, 2003). A clue to the contextual complexity of TGF-β signalling emerged in a recent paper demonstrating that master transcription factors direct Smad3 binding to their respective target loci (Mullen et al., 2011). Genome-wide Smad3 binding sites were assessed in ES cells, pro-B cells and myotubes, identifying that SMAD3 bound regions are almost entirely unique in each cell type, with only 1% overlap. Smad3 co-localized to over 70% of genomic loci with the master transcription factors OCT4, SOX2 and NAOG in ES cells, PU.1 in pro-B cells and MYOD1 in myotubes. Upon TGF-β stimulation or inhibition, only the genes that were bound by SMAD3 and master transcription factor were differentially regulated in the respective cell type. Interestingly, over-expression of *MyoD1* in ES cells was able to recruit SMAD3 to MYOD1 binding sites in ES cells, while Smad3 also remained bound at pluripotency loci. Those findings nicely

demonstrated that master transcription factors recruit SMAD3 to their target loci and that the transcriptional output of the system depends on levels of TGF-β/Smad3 activity.

The aforementioned examples of diverse Smad2/3 actions in unique cellular contexts demonstrate that Smad's serve as cofactors potentiating transcriptional regulation at genomic regions as specified by master transcription factors. In the context of cell identity changes, such as those occurring during development, it is crucial to have tight control of timing and amplitude of gene expression. With the ability to modulate Smad signalling through TGF-β ligand gradients, such a system would impart cells the ability to tightly regulate master transcription factor signalling output. In the context of forced cell identity changes with master transcription factors, such as the reprogramming of somatic cells to pluripotency, it has been previously demonstrated that tight regulation of TGF-β signalling is instrumental to successful reprogramming. Yet, little is understood as to the roles of TGF-β signalling throughout the process.

1.5 Aims of this thesis

The aim of this work is to characterize the contributions of TGF-β signalling during reprogramming to pluripotency. Secondarily, the work will test the hypothesis that discoveries of TGF-β function during reprogramming might benefit other forced cell identity conversions with master transcription factors. When considering the large cohort of identified TGF-β target genes, there may be a multitude of mechanism through which these signalling pathways alter the dynamics of cell identity conversions. While many publications have suggested that TGF-β inhibition during fibroblasts reprogramming to iPSCs blocks mesenchymal signalling, promoting the acquisition of epithelial character, most works have not distinguished if such observations are the cause or consequence. Through a combination of reprogramming surface marker expression dynamics, a *Nanog*-GFP pluripotency reporter, monitoring SMAD2/3 activity, and over-expression of constitutive active *Smad2/3* constructs, these works explore the contributions of TGF-β signalling during reprogramming and apply those findings to other cell identity conversion systems.

1.6 A brief overview of findings from this work

The first objective was to monitor the mesenchymal to epithelial transition during reprogramming, revealing that the TGF-β receptor inhibitor (Alki) did not accelerate the up-regulation of *E-cadherin*. FACS analysis and sorting with reprogramming stage-specific surface markers CD44 and ICAM1 revealed that the Alki boosts intermediate stage transitions during reprogramming, long after the mesenchymal to epithelial transition occurs. To explore the downstream mechanism of Alki action, the canonical TGF-β pathway transcription factor Smad3 was targeted with a specific inhibitor of Smad3 (SIS3). Unexpectedly, the SIS3 was unable to recapitulate the Alki results; SIS3 almost completely inhibited reprogramming. Following those results, the activity of SMAD3 during reprogramming was monitored by immunofluorescence with a phophorylated-SMAD3 (pSMAD3) specific antibody, which recognizes the active form of SMAD3. Unexpectedly, pSMAD3 levels were higher in the presence of Alki during reprogramming. Furthermore, cells with high levels of pSMAD3 were also expressing the highest levels of *Nanog*-GFP reporter. Those results suggested that the level of active SMAD3 is somehow increased in the presence of Alki, perhaps due to a skewed negative feedback loop, which might be crucial for enhanced reprogramming by Alki. This hypothesis lead to an experiment whereby a constitutive active *Smad3* (*Smad3CA*), or *Smad2CA*, was over-expressed during reprogramming. Strikingly, *Smad3CA* or *Smad2CA* drastically boosted the efficiency of reprogramming over 6-fold, driving *Nanog*-GFP expression and substantially accelerating the reprogramming process. Global RNA-seq analysis confirmed that reprogramming intermediate stage-specific transient gene up-regulation is less evident in the presence of SMAD3CA, indicating SMAD3CA drives cells toward a pluripotent state on a more direct route. Co-immunoprecipitation studies then revealed interactions between SMAD3 and the reprogramming master transcription factors OCT4, SOX2 and to a lesser extent KLF4. SMAD3 also interacted with a multitude of chromatin remodellers and transcriptional activators during reprogramming. Finally, because SMAD2/3 are well characterized to interact with cell type specific master transcription factors, transdifferentiation systems were investigated to explore if *Smad3CA* could boost other cell identity conversions in

collaboration with Thomas Graf's and Malin Parmar's laboratories. Indeed, expression of *Smad3CA* boosted the conversion of B cells to Macrophages, myoblasts to adipocytes, and fibroblasts to neurons. These results reveal Smad2/3 as potent amplifiers of cell identity conversions with master transcription factors and identify a powerful tool to boost these processes.

2: Materials and Methods

2.1 Cell culture

2.1.1 Cell culture reagents

ES medium:

- Glasgow Minimal Essential Medium (GMEM, Sigma G5154)
- Fetal Calf Serum (FCS) (10%) (Life Technologies)
- Non essential amino acids (1x, Gibco 11140-035)
- L-Glutamine (2 mM, Invitrogen)
- Sodium pyruvate (Invitrogen)
- 100 U/ml LIF (human recombinant, made in house)
- Penicillin/Streptomycin (100 units/100 ug, Sigma P4333)

MEF medium:

- ES media
- FGF2 (5 ng/mL, Peprotech, 100-18-B)
- Heparin (1 ng/mL, Sigma)

Reprogramming media:

- ES media
- Doxycycline (300 ng/ml for reprogramming of targeted cell lines or 1000 ng/ml for primary reprogramming) (Sigma)
- Ascorbic Acid (VitaminC) (10 ug/mL, Sigma 1000731348)

- +/- ALKi (A83-01) (500 nM, TOCRIS Bioscience #2934)
- +/- SB43 (SB431542) (10 uM, TOCRIS Bioscience, #1614)

PlatE media:

- ES media (10 μg/ml)
- Puromycin (1 μg /ml)
- Blasticidin (10 μg /ml)

Freezing solution:

- DMSO (10%, VWR International)
- FCS (90%) (Life Technologies)

Other reagents:

- Gelatin (0.1% in PBS, Sigma G5154)
- Trypsin (0.25%, Gibco 15090-046)
- EDTA (0.1%, Sigma 03620)
- PBS (Sigma, D8537)
- Accutase (Sigma, A6964)

2.1.2 Cell lines used

- TNG-MKOS
- 129 or CD1 derived MEF
- PlatE viral packaging line

The cell line used for reprogramming, termed TNG-MKOS, has been described in a recent publication (Chantzoura E.; Skylaki, in press). Briefly, a mouse ESC line (TNG ESC line) harbouring a targeted *Nanog*-GFP (TNG) reporter with GFP coding sequence at the first ATG site of the *Nanog* exon 1 was provided by Dr. Chambers. With

this cell line, expression of GFP reports the endogenous *Nanog* promoter activity, while also producing a *Nanog* heterozygous mutant (Chambers, 2007). Furthermore, the GFP cassette in the *Nanog* locus has an IRES-*puromycin* resistance gene cassette, enabling selection of *Nanog*-GFP expressing cells. A doxycycline inducible polycystronic cassette with the 4 reprogramming factors *c-Myc (M), Klf4 (K), Oct3/4 (O), Sox2 (S*) linked by 2A peptides, followed by an internal ribosome entry site (IRES) mOrange reporter and a CAG promoter driven-rtTA (doxycycline transactivator) expression cassette was integrated into the TNG ESC line in to the Sp3 locus via gene targeting. The Sp3 locus had been identified to be conducive to reprogramming from random integration of piggyBac transposon carrying a dox-inducible reprogramming cassette previously (O'Malley et al., 2013).

The wild type (wt) MEFs used for reprogramming experiments were derived from the mouse strain 129 or CD1. MEF were harvested as described below.

The PlatE viral packaging system was used for retroviral mediated gene delivery. The platE line was previously created by transfecting the 293T cell line with EF1α promoter driven cassettes expressing the *gag-pol* and *env* viral packaging genes followed by IRES-*blasticidin/puromycin* resistance sequences to enable maintenance of high titer lines (Morita et al., 2000). Viral mediated gene transduction is described in more detail below.

2.1.3 Cell passaging

Both ES and MEF were cultured until ~80-90% confluent at which point they were passaged at a ratio of 1:5 or 1:3 respectively. Passage was performed by 1 wash with PBS, followed by 3 minutes treatment with Trypsin (0.25%) EDTA (2 mM) then re-suspended in medium and centrifuged at 300 g for 3 minutes prior to re-suspension in their specific media and plating.

2.2 Reprogramming experiments

2.2.1 Generation of reprogrammable MEF

The TNG-MKOS ESC line was used to produce chimeric embryos by morula aggregation. Prior to morula aggregation, cells were plated at 2X serial dilutions starting from $2x10^6$ cells down to $62.5x10^5$ cells. The next day, colonies of 5-8 cells were then selected for aggregation with morulas (2.5 d.p.c) of C57B1/6 mice. The staff at our in-house transgenics facility carried out the procedures, as previously described (Pluck and Klasen, 2009). The implanted chimeric embryos from morula aggregation were harvested at E12.5 for MEF isolation (described below) and reprogramming.

To generate reprogrammable transgenic mouse lines, we performed blastocyst injection with the TNG-MKOS ESC line. The resulting chimeric mice (recipient C57B1/6 + ESC 129 background) were crossed with CD1 mice. We chose CD1 mice because they breed well, different from 129 mice, and the maintenance of CD1 ESC lines is as easy as that of 129, easier than from that of C57B1/6. $Nanog^{+/GFP}$,$Sp3^{+/tetO\text{-}MKOS,\ CAG\text{-}rtTA}$ F1 male mice were used as breeders. One quarter of the pups from the F1 x wt CD1 breeding were the desired genotype as we expected from the Mendelian inheritance ratio.

2.2.2 Extraction of MEF for reprogramming

Embryos were dissected from culled mothers at E12.5 for harvesting of MEF. The embryos head was removed (saved for genotyping) and the body was eviscerated and placed in 0.25% trypsin EDTA for 10 minutes at 37^oC. They were further dissociated by passage through an 18-gauge needle 10 times, prior to another 5 minutes incubation at 37^oC. Trypsin was inactivated with MEF media, then cells were centrifuged and re-suspended in MEF media and plated into a 10 cm plate; 1 embryo per plate. MEF were expanded for 2-3 days prior to freezing them at $2x10^6$ cells per vial.

2.2.3 Genotyping embryos

The head or endoderm of embryos was added to 25 µl of lysis buffer containing 25mM NaOH, 0.2 mM EDTA and incubated at 95° for 20 minutes, then cooled to 4 °C. Then 25 µl of neutralization buffer containing 40 mM Tris HCl (pH 5.0) was added and the sample was vortex mixed for 10 seconds. The final solution contains 20 mM Tris/0.1 mM EDTA/pH 8.0 with 12.5 mM sodium ions. For PCR reaction, 1 ul of lysate DNA solution was used for a 15 ul PCR reaction containing: 1 ul DNA solution, 1 µl each of 2 µM primers, 0.075 ul DreamTaq polymerase (Thermo Scientific), 7.5 µL 2X buffer, 2µL dNTPs, 2.425 ddH_2O. PCR was run with primers specific to one or all of Nanog/GFP/rtTA (primer table below) as necessary, using standard PCR protocol involving a pre-incubation at 95^oC for 3 minutes, 20 Cycles of 95^oC/60^oC/72^oC for 10s/5s/30s respectively, followed by a 5 minute incubation at 72^oC and storage at 4^oC indefinitely.

Name	**Sequence**	**Genotype**	**Amplicon size (bp)**
N-Ex1 F	CTGAGGAAGCATCGAATTCTGG	Nanog wt allele	186
N-Int1 R	AAGCAATAACCCTTCAGCCC	Nanog wt allele	186
IC346 F	ATATGGCCACAACCATGACC	Nanog-GFPiPuro	211
IC347 R	ACCCACACCTTGCCGATGTC	Nanog-GFPiPuro	211
mO F1	ACCATGGTGAGCAAGGGCGA	MKOSimOrange	241
mO R1	GCTTCACGTAGGCCTTGGAG	MKOSimOrange	241

Table 2.1. Primers for genotyping

A list of primers for genotyping MEF for reprogramming

2.2.4 TNG-MKOS MEF reprogramming

I used 2 general strategies for reprogramming experiments, which involve a starting population of either 5% transgenic reprogrammable MEF amongst 95% wt feeder MEF, or 100% transgenic MEF to start. The majority of experiments in this work are from the 5% starting transgenic cell experiments; conditions which were previously established as optimal for our reprogramming system. However, some experiments require a large quantity of cells undergoing reprogramming without contamination of wt MEFs, such as those for ChIP-qPCR. In these cases, I used 100% transgenic cells to avoid requirement of cell sorting. Whether starting with 5% or 100% transgenic cells, the reprogramming conditions remain the same.

MEF are defrosted at high confluence (>70%) in MEF media 2-3 days prior to plating for reprogramming. For 5% transgenic cell reprogramming, one day prior to initiating reprogramming, the cells are seeded on gelatin pre-coated plates at $1x10^5$ total cells for 1 well of a 6-well plate, or $7x10^5$ total cells for a 10cm plate, which is $5x10^4$ or $3.6x10^5$ transgenic cells respectively. For 100% transgenic cell reprogramming experiments, $5x10^4$ or $5x10^5$ transgenic cells are plated in 1 well of a 6 well or 10 cm dish respectively. One day after plating, MEF media is replaced with reprogramming media containing doxycycline (dox) to activate expression of MKOS reprogramming factors. Reprogramming media is replenished every 2 days until the end of the experiment.

2.2.5 Retroviral mediated gene transduction

Viruses can be exploited as gene delivery tools to package, transduce and express genes of interest into the mammalian genome (Cherry et al., 2000; McTaggart and Al-Rubeai, 2002; Vu et al., 2008). In these studies we utilized the pMXs retroviral gene delivery system (Kitamura et al., 2003). Retroviral systems deliver viral RNA into the host cell nucleus, which then gets reverse transcribed by a viral packaged reverse-transcriptase into a DNA template and is incorporated into the host genome via long terminal repeats (LTRs), aided by the viral integrase (McTaggart and Al-Rubeai, 2002).

Upon genome insertion, the 5' LTR drives transcription of the viral DNA insert. For introducing a gene of interest into host cells, one can clone any gene into the viral genome plasmid between the LTR sites, which, upon viral production and transduction, will be incorporated into the host genome and be transcribed (Kitamura et al., 2003; McTaggart and Al-Rubeai, 2002). For safety and efficacy, all of the necessary genes encoding viral packaging proteins were stably transfected into a viral packaging cell line called PlatE, such that viral replication genes are not packaged into the viral particles. The PlatE cell line constitutively expresses the *gag-pol* and *env* packaging genes, allowing for production of a virus coding for integration and expression of a gene of interest with transfection of a single plasmid, that cannot produce more virus in the infected cell (Morita et al., 2000).

For reprogramming experiments, the pMXs system was used because it exhibits far less toxicity in MEF as compared to the highly infectious lentiviral packaging systems (observations from our lab). The pMXs retroviral expression system was used to over-express genes of interest (GOI) in our TNG-MKOS reprogramming system. For retroviral production, the PlatE cell line was grown to ~80% confluence, and then $1.6x10^6$ cells were plated in a 10 cm dish in 8 ml of media. One day later the cells were transfected with a pMx viral vector coding for expression of a gene of interest (GOI) (see plasmid list). Transfection was carried out using the $CaCl_2$ based delivery method (Kingston et al., 2003). Briefly, 10 μg of pMXs-GOI DNA was made to a final volume of 427 ul in H_2O, then added to 63 ul 2M $CaCl_2$ and mixed well; the DNA-$CaCl_2$ mixture is added into 500 ul 2X HBS (8 g NaCl, 0.2 g Na_2HPO_4-$7H_2O$, 6.5 g HEPES in 500 ml H_2O, pH 7) by pipetting repeatedly, and then 800 μl of the final mixture was added to the PlatE in 8 ml of media. After 24 hours, the transfected PlatE cell medium was changed, and virus was allowed to accumulate in the media for 24 hours, at which point viral infection was initiated. Viral infection is initiated by first filtering the virus containing supernatant through 0.45 μM Whatman cellulose acetate filters to remove contaminating PlatE cells and then adding polybrene (Merck-Millipore) at 8 μg/ml to the supernatant to enhance viral particle absorption in MEF (Davis et al., 2002). The supernatant is then added to the MEF, 2ml in a well of a 6-well, or 8 ml into a 10 cm

plate. The MEF were exposed to viral supernatant for 6 hours at 32 ^{o}C (7.5% CO_2), and then media is changed to reprogramming media and plates are placed back at 37 ^{o}C (7.5% CO_2). All virus procedures were performed in a designated retroviral hood, according to established lab viral protocols.

2.2.6 Generation of Smad2CA/3CA with site directed mutagenesis

To create the Smad2CA/3CA expression constructs, the cDNA of the genes was cloned using primers that introduce base pair changes in the genes. For Smad2CA/3CA expression constructs, PCR primers were designed with base pair mismatches at the 3' end, to exchange serine codons for aspartic acid (SSVS→DDVD*) in *Smad3*, or glutamic acid (SSMS→SEME) in *Smad2*, in the C-terminus of the proteins (Chipuk et al., 2002; Funaba and Mathews, 2000). The primers used for mouse Smad2 were forward- CTAGGGTAGATTTACCGGGC, Reverse-CGAGTCTTTGATGGGTTTACTCCATCTCTGAGCATCGCACTGAA, and for Smad3 forward-GCTGGCGCCGGAACCAATTCAGTCGACGTGACCCTTCGGTGCCAG, reverse-CTAATCCACATCGTCACAGCGGATGCTCGG GGAACCCATCTGGGT. Primers for human Smad2CA/3CA cloning followed a similar strategy with Smad2CA forward-CCGGAACCAATTCAGTCGACTGGAATTCGCCACCATGTCGTCCATCTTGCCA TTCAC, reverse-TCTCGAGTGCGGCCGCGAATTGGATCCTTAGTCCATGTCTTCGCAACGCACT GAAGGGGAT, and Smad3CA forward-CCGGAACCAATTCAGTCGACTGGAATTCGCCACCATGTCGTCCATCCTGCCT TTCA reverse-TCTCGAGTGCGGCCGCGAATTGGATCCCTAATCCACATCGTCACAGCGGATG CTTGGGGA. The PCR products were then cloned into the TOPO Blunt vector and subsequently cloned into expression vectors.

2.2.7 Sorting and re-plating reprogramming experiments

For some experiments, cells were sorted and re-plated for continued reprogramming. In such experiments, cells were sorted into FACS buffer (2% FCS in PBS) and then re-plated onto gelatin coated 6-well plates in reprogramming media. They were seeded as $5x10^3$ or $1x10^4$ cells into 1 well of a 6-well, depending on stage of reprogramming. Cells were reprogrammed for an additional 10 days, prior to Nanog-GFP^+ colony counting.

2.2.8 Plasmids for reprogramming

Associated experiment	Plasmid name	List of features	Promoter	Resistance
TNG-MKOS reprogramming-retroviral top-up	pMXs-Smad3CA	LTR-Smad3CA	LTR	Ampicillin
TNG-MKOS reprogramming-retroviral top-up	pMXs-Smad2CA	LTR-Smad2CA	LTR	Ampicillin
TNG-MKOS reprogramming-retroviral top-up, control vector	pMXs-BFP	LTR-BFP	LTR	Ampicillin
TNG-MKOS reprogramming-retroviral top-up, control vector	pMXs-dsRed	LTR-BFP	LTR	Ampicillin
Primary reprogramming	CAG-MKOS	CAG-MKOS	CAG	Ampicillin
Primary reprogramming	TET-MKOSimO	TET-MKOS-IRES-	TET	Ampicillin

		mOrange		
Primary reprogramming	TET-Smad3CAimO	TET-Smad3CA-IRES-mOrange	TET	Ampicillin
Primary reprogramming	CAG-rtTAiH	pyCAG-rtTA-IRES-Hyg^{R}	pyCAG (Fimia et al., 1995)	Ampicillin
Primary reprogramming	hyPBase	CMV-hyPBase (Yusa et al., 2011)	CMV	Ampicillin

Table 2.2. Plasmids for reprogramming

A list of plasmids used in various reprogramming experiments.

2.3 Florescence Activated Cell Sorting (FACS)

Antibodies:

- CD44-biotin conjugate (1:100 dilution) (eBioscience 13-0441)
- CD44-APC conjugate (1:300 dilution) (eBioscience 17-0441)
- Icam-1/CD54-biotin conjugate (1:100 dilution) (eBioscience 13-0541)
- E-cadherin-eFluor660 conjugate (1:300 dilution) (eBioscience 50-3249-82)
- Strepdavidin PE-Cy7 (1:1500 dilution) (eBioscience 25-4317)

2.3.1 Cell preparation for FACS

Cells were lifted with Trypsin-EDTA or Accutase (for E-cadherin experiments) for 5 or 15 minutes respectively. Trypsin/accutase was inactivated in serum and then

cells were spun and re-suspended in FACS buffer (2% FCS in PBS). Cells were washed twice in FACS buffer, re-suspended at ~1x10^6 cells per 300 µl, prior to staining with primary antibodies, on ice for 10 minutes. Cells were centrifuged at 300g for 3 minutes prior to 3 washes in PBS and then re-suspended to 1x10^6 cells in FACS buffer with secondary antibodies and incubated on ice for 5 minutes. Stained cells were then washed 3 times in PBS, filtered in a 40 µM mesh filter (BD) prior to re-suspension in FACS buffer at original volume and then processed using FACS analysis or sorting. FACS sorting or analysis were carried out on the BD FACS AriaII or Fortessa respectively, according to conditions as outlined in (O'Malley et al., 2013), and summarized below. Laser voltages varied and were adjusted each time according to controls.

2.3.2 FACS settings

	Laser		
Band filter	488nm	561nm	640nm
530±30	GFP		
582±15		mOrange	
670±30			APC
780±60		PE-Cy7	

Table 2.3. FACS settings

Laser refers to the wavelength of light the fluorophore bound sample was excited with. The band filter indicates the wavelength of light detected by the instrument.

2.4 Immunofluorescence Microscopy

2.4.1 Antibodies

- p19Ink4d (Abcam)(1:100)

- p16Ink4a (Santa Cruz Sc-1661)(1:100)
- pSmad3 (C25A9 Cell Signalling) (1:100)
- Smad3 (ab28379, Abcam)(1:100)
- anti-mouse/rabbit/strepdavidin AlexaFluor488/594/647/750/Pacific Blue (life technologies)(1:1000)
- Nanog (Abcam, #Ab14959)(1:300)
- See FACS antibodies

2.4.2 Sample preparation

Cells were reprogrammed in 6 wells on 18 mm circular cover slips (Fisher Scientific) or directly on 6 well Corning plates for colony counting experiments. Cells were fixed in 4% paraformaldehyde (PFA) in PBS for 10 minutes, then permeabilized in 0.1% Triton in PBS for 1-hour at room temperature. Samples were then blocked in 5% normal goat serum in PBS for 1 hour, prior to incubation with primary antibodies in blocking solution at 4 ^{o}C overnight. The next day cells were washed 3 times with PBS, and then incubated with secondary antibodies in blocking solution for 1 hour at room temperature, followed by 3 washes with PBS. Samples in 6 wells were directly imaged, whereas samples on cover slips were mounted on slides (Fisher scientific) using Prolong Gold (+/- Dapi) and then imaged. For confocal microscopy, all imaging was performed on a Leica TSC SP2 and processed using Adobe Phooshop. For entire well colony and cell counting, the images were captured and analyzed using the *Celigo* S Cell Cytometer (Nexcelom). All image processing, including adjustment of gain and range of signal, was performed identically for each channel across all samples in any given figure.

2.5 Western Blotting

Cells were harvested using Trypsin-EDTA and then washed once with PBS prior to lysis with 1X NuPAGE LDS Sample buffer (Life Technologies) (100-250 µl) with phosphatase inhibitors (HALT, Life techologies). Lysates were heated to 95^{o}C for 10

minutes, followed by sonication of 3 cycles of 15 seconds on a misonix XL2000 sonicator on setting ~2. Protein concentration was quantified using a Pierce BCA Protein Assay Kit (Life technologies) and BSA standard curve. From 1-100 μg lysate was mixed with DTT (final concentration 100 mM) and then run in NuPAGE MOPS running buffer at 130V for 1.5 hrs on a NuPAGE 10-12% Bis Tris Gel (Life Technologies) using a BioRad Mini-PROTEAN electrophoresis machine. Proteins were then transferred to 0.45 μM Protran nitrocellulose membrane (Sigma Aldrich) on ice at 50V in transfer buffer (25 nM Tris, 190 mM glycine, 20% Methanol, pH8.3) for 2-4 hours with a Mini Trans-Blot Cell (Bio-Rad). The membrane was then blocked in 5% Skim Milk (Sigma), or 5% BSA (for phosphorylated proteins), in TPBS (0.05% Tween in 4000 ml dH_20 containing 121.14 g of Tris base and 438.3 g of NaCl). After blocking, membrane was incubated over night at 4 ^oC with primary antibody in blocking buffer. The following day the membrane was washed 3 times with TBST, incubated with HRP conjugated secondary antibody for 1 hour, washed 3 times with TBST, and then imaged using the ECL based SuperSignal West Pico Chemiluminescent Substrate with X-ray film (Kodak, Biomax) and a Konica SRX101A developer.

2.6 Co-immunoprecipitation

Co-immunoprecipitation using nuclear extract lysate was adapted from a previously established protocol (Mullen et al., 2011). Samples were first incubated in 0.5% NP40 in hypotonic lysis buffer (10mM HEPES, 1.5 mM MgCl2, 10 mM KCl) with proteinase (cOmplete protease inhibitor cocktail, Roche) and phosphatase inhibitors (HALT, Life Technologies), vortexed and left on ice for 5 minutes prior to centrifugation at 500 g to pellet nuclei. The pellet was suspended in high salt solution (HEPES pH7.9, 1.5mM $MgCl_2$, 0.2mM EDTA, 400mM NaCl, 25% Glycerol) to extract nuclear protein. As a lysate pre-clearing step, Dynabeads-ProteinG (Dynaneads ProteinG, Life Technologies, #10003D) were blocked for 1 hour in 2% skim milk in no salt-Buffer (20 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 20% Glycerol, phosphatase and protease inhibitors), prior to incubation with the nuclear isolated samples overnight in IP-buffer (20 mM Hepes, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 20%

Glycerol, 100 mM NaCl, protease and phosphates inhibitors). 2 ug of antibody or IgG was added to 25 uL blocked Dynabeads for an overnight incubation in IP-buffer. Antibody bound beads were then magnetically isolated and added to 200 ug pre-cleared lysate for a 3-hour incubation, rotating at 4^{o}C. Beads were then washed 3 times in IP buffer and processed for western in LDS buffer as describe above. Antibodies included Rb-Smad3 (Abcam, #ab28379), ms-Smad3 (Abcam, #AF9F7), Oct4 (Santa Cruz Biotechnology, #sc8628), Sox2 (Santa Cruz Biotechnology, #sc17320), Klf4 (Santa Cruz Biotechnology, #sc20691), Dpy30 (Santa Cruz Biotechnology, #sc167677), p300 (Thermo Scientific, #MS-586-PO), Brg1 (Santa Cruz Biotechnology, #sc17796), Med15 (Sigma Aldrich, SAB2500761), and IgG (Millipore).

2.7 Chromatin Immunoprecipitation followed by quantitative PCR

Chromatin immunoprecipitation (ChIP) experiments were executed in 3 stages: first the cells are fixed in formaldehyde and sonicated, then they are precipitated with antibody bound beads, and finally the immunoprecipitated DNA is analyzed using quantitative PCR (qPCR).

2.7.1 Chromatin fixation

Cells were grown to 80% confluence in 10 cm dishes, and then fixed for 12 minutes on a shaker in formaldehyde at 1% final concentration. Formaldehyde was quenched by addition of 1 M glycine (0.125 M final concentration) for 5 minutes shaking at room temp. Samples were then washed once in PBS (with protease inhibitors), collected with silicon scrapers and centrifuged at 300 g for 5 minutes. Cells were re-suspended in PBS with protease inhibitors and aliquot in 10^7 cell batches (based on counts from a sacrificed plate) into eppendorf tubes, centrifuged at 300 g, supernatant was removed and pellets were snap frozen on dry ice.

2.7.2 Nuclear isolation and sonication

A pellet of 10^7 cells was thawed and re-suspended in lysis buffer (5 mM Pipes pH 8, 85 mM KCl, 0.5% NP40) with fresh protease inhibitors (cOmplete protease inhibitor cocktail, Roche) and incubated on ice for 20 minutes. Nuclei were subsequently pelleted by centrifugation at 500 g for 10 minutes at 4 ^{o}C. Nuclear pellets were re-suspended in IP buffer (0.5% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris-HCl ph 8, 150 mM NaCl, fresh protease inhibitors), and sonicated in eppendorf tubes as follows: 5-8 cycles of 30 second pulses on a BioRuptor Sonicator, on the "high" setting, each cycle consisting of 10 minutes of 30 seconds on, followed by 30 seconds off. A typical sample of 10^7 cells from a reprogramming experiment required 8 cycles of 10 minutes on/off 30 second intervals, over the coarse of 80 minutes. Such conditions produced samples with an average band size of ~150 bp, as assessed by DNA gel electrophoresis (shown below). After sonication, samples were diluted 2X in IP buffer without SDS, to produce a final SDS concentration of 0.25%. Then, 8ug antibody or IgG was added to samples for overnight incubation, rotating at 4 ^{o}C.

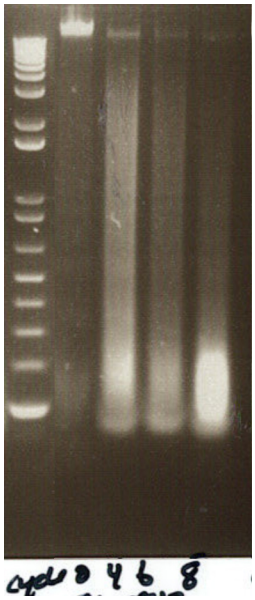

Figure 2.1. Sonicated DNA

An image of an agarose gel with SybrSafe (25000X) DNA stain to visualize examples of fixed and sonicated DNA from 10^7 cells used for ChIP experiments. The ladder is a 1kb plus ladder (Life technologies), whereby the 3 bottom bands represent 100, 200 and 300 bp, from bottom to top respectively. After the ladder, lanes 1-4 represent 0, 4, 6 and 8 cycles of sonication respectively. Sonication for 8 cycles was often used for reprogramming experiments, to attain average fragment size of ~150 bp.

2.7.3 Immunoprecipitation and DNA isolation

60 ul of Dynabeads-ProteinG (Dynaneads ProteinG, Life Technologies, #10003D) were blocked in 0.5% BSA in IP buffer for 1-hour, rotating at 4^{o}C. Blocked beads were then magnet extracted and the antibody bound chromatin suspensions were added to the pre-blocked bead pellet and rotated at 4^{o}C for 4 hours. Bead-antibody-chromatin complexes were then magnet isolated, followed by a series of washes with increasing stringency to remove non-specific binding of Beads to proteins or DNA. Each wash was 5 minutes, rotating at room temperature, followed by magnet bead extraction and re-suspension in the next wash buffer. The first wash was with IP buffer (150mM NaCl), followed by IP buffer with high salt concentration (500mM NaCL), then 1 wash with washing buffer (10mM Tris-HCl ph8, 0.25M LiCl, 0.5% NP40, 0.5% Na-Deoxycholate, 1mM EDTA), followed by 2 washes with TE (10mM Tris, 1mM EDTA). The DNA-protein complexes were eluted from the beads by incubation in 100uL elution buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl pH8) for 15 minutes at 65^{o}C. Beads were magnet extracted and DNA containing supernatant collected in a new 1.5ml eppendorf tube. Beads were further washed with 150ul TE (with 1% SDS), magnet extracted and the suspension was added to the 100uL elution buffer containing DNA. The samples were incubated at 65^{o}C overnight to reverse crosslink the DNA-protein complexes. The following day, the ChIP'd DNA was treated for 2 hours with 5ul (4 units) proteinase K (NEB) at 37^{o}C.

DNA was then isolated using Sera-Mag beads as described below, using 2.5X bead:DNA suspension ratio (12.8% PEG final concentration). After DNA isolation, beads were re-suspended in 30uL ddH_2O, magnet extracted, and DNA solution was treated with 2ul RNAse A (Life Technologies) at 37^oC for 20 minutes and then stored at -20^oC. For qPCR, 0.5µl DNA was used per reaction.

2.8 RNA sequencing

All RNA sequencing was performed in collaboration with Anna Johnsson and Sten Linnarsson at the Karolinska Institute in Stockholm, Sweden, according to their published works (Islam et al., 2011; Islam et al., 2014). The samples were collected by FACS, then 2500 cells were aliquot (500 cells/ul) into PCR strip tubes in PBS containing RNAse inhibitor and snap frozen on dry ice prior to shipment to Sweden. The cDNA preparation and sequencing was performed as published (Islam et al., 2011; Islam et al., 2014). Briefly, universal oligo-dT probes are used to reverse transcribe all mRNA. As the MMLV reverse transcriptase reaches the 5' end of mRNA, it generally adds a few oligonucleotides (predominantly cytosines) to the 5' end of the transcript. Thus, this oligoC addition to the endogenous mRNA allows for addition of a barcode on the 5' end of all transcripts through incorporation of barcoded DNA sequence containing oligo-GGG on its 3' end (6bp Barcode-GGG). The barcoded template hybridizes to the 5' end of the transcript and as the MMLV encounters this hybrid, it undergoes a template switch, and proceeds to extend the newly synthesized cDNA to include the barcode. Thus, each prepared sample has a unique barcode, which allows pooling of samples for all subsequent library prep, producing a more uniform, unbiased pool of libraries. Recent optimization of the protocol exploits this 5'-CCC strategy to introduce unique molecular labels (UMLs), reliant on a large pool of 6-bp barcodes, that when added at random to the transcripts prior to amplification, allows for each original transcript to be recorded, thereby allowing for a more absolute quantification of each transcript by means of molecule counting after sequencing (Islam et al., 2014). After reverse transcription and UML incorporation, the full-length cDNA is fragmented and further bar-coded with adapters for PCR amplification, in a step known as Tn5

transposon mediated fragmentation, or tagmentation. Amplified libraries are then sequenced using an Illumina HiSeq sequencing platform.

2.9 RNAseq data analysis

Jon Manning, a resident informatician at our institute, performed all data analysis. RNA-seq was performed with the single-cell tagged reverse transcription (STRT) platform using 500 cells/sample, which had been snap frozen in PBS with RNase inhibitor, then processed as previously described (Islam et al., 2012; Islam et al., 2014). The RNAseq reads were processed by an automated pipeline as outlined in a recent publication (Islam et al., 2014). Counts data, based on both read and molecule counts, were processed with the DESeq2 (Version 1.6.3) package of Bioconductor (Love et al., 2014). However we found that the processing steps used to produce molecule count data lead to the loss of read counts for low-abundance genes, so proceeded with analysis on read count data, analogous to standard RNA-seq. 2-3 technical replicates per sample tightly clustered and therefore summed across columns prior to further data processing. As per the standard DESeq2 protocol, normalized expression estimates were obtained by adjusting columns by a size factor corresponding to library size. For read counts, for example, this adjustment ranged from 0.57 to 0.71. Data were transformed to log2 scale by use of DESeq2's rlog() command, which also minimizes differences at very low count levels. The clustering patterns of genes in the two series were assessed based on a matrix of the mean of biological replicate samples. Two matrices were constructed containing the Cont and +SmCA3 series. The mean of MEF and ES samples were included in both series. To reduce noise, genes with background level expression (no mean count >= 2)

or low variation (coefficient of variation < 100) in both series were removed from this matrix. The matrix was clustered by use of the clara() function in R, which approximates the Partitioning About Medoids algorithm. Mean normalized counts of each cross-classified gene group (with >100 genes) identified in the chord diagram were shown.

2.10 Quantitative polymerase chain reaction (qPCR)

The qPCR reactions were performed in 384 well plates, in a 10ul reaction, using the LightCycler480 instrument. For a typical qPCR reaction, 10ng of DNA was loaded to each well. Depending on experiment, either the SybrGreen (Life Technologies) or the universal probe library (UPL) (Roche) system of detection was used, according to manufacturers protocol. For each reaction, 0.5uM primers were used.

2.10.1 qPCR primers table

Name	Sequence	UPL probe	Experiment
p19ARF-F	GGGTTTTCTTGGTGAAGTTCG	106	Senescence
p19ARF-R	TTGCCCATCATCATCACCT	106	Senescence
p16Ink4a-F	AATCTCCGCGAGGAAAGC	91	Senescence
p16Ink4a-R	GTCTGCAGCGGACTCCAT	91	Senescence
mouseSmad3-F	TCCGTATGAGCTTCGTCAAA	32	Time course reprogramming
mouseSmad3-R	GGTGCTGGTCACTGTCTGTC	32	Time course reprogramming
Nanog-F	GCCTCCAGCAGATGCAAG	25	Time course

			reprogramming
Nanog-R	GGTTTTGAAACCAGGTCTTAACC	25	Time course reprogramming
Upstream Oct4-F	AGCCTTTTGAAATTCAGGAGCAG	NA	ChIP-qPCR Smad/Oct Binding
Upstream Oct4-R	AAAAAGAAACCCTGACTAGACACCA	NA	ChIP-qPCR Smad/Oct Binding
Upstream Nanog-F	CAGTGAGTTCCAGGCTAACCA	NA	ChIP-qPCR Smad/Oct Binding
Upstream Nanog-R	ACTCGGACTAACCAAGGGCTA	NA	ChIP-qPCR Smad/Oct Binding
Sox2-Smad3-F1	AAGCTAGGCAGGTTCCCCTC	NA	ChIP-qPCR Smad/Oct Binding
Nanog-Smad3-F1	TCAGATCCCCCACTTGACCT	NA	ChIP-qPCR Smad/Oct Binding
Pou5f1-Smad3-F1	GGAGGTTGAGAGTTCTGGGC	NA	ChIP-qPCR Smad/Oct Binding
Klf2-Smad3-F1	GGGGATGTGCACCTGATGAA	NA	ChIP-qPCR Smad/Oct Binding
Sall1-Smad3-F1	CTTCGGCTGAGTAATGGGGG	NA	ChIP-qPCR Smad/Oct

			Binding
Sall4-Smad3-F1	TTAAAAGCGGCGCCACTAGA	NA	ChIP-qPCR Smad/Oct Binding
Fbxo15-Smad3-F1	CCGCCCTTAGTTCCCAGATG	NA	ChIP-qPCR Smad/Oct Binding
Gdf3-Smad3-F1	CAGCAGGGGGATTGCTGTAA	NA	ChIP-qPCR Smad/Oct Binding
Rex1-Smad3-F1	CTGTCCCCTTGTCTTGGGTC	NA	ChIP-qPCR Smad/Oct Binding
Jarid2-Smad3-F1	GGCTGCTCAGTCTTCCTGTC	NA	ChIP-qPCR Smad/Oct Binding
Dppa2-Smad3-F1	GGACAACCTGGGCTTTGTCT	NA	ChIP-qPCR Smad/Oct Binding
Esrrb-Smad3-F1	TAGAACCCACAACGTGACCG	NA	ChIP-qPCR Smad/Oct Binding
Dppa4-Smad3-F1	GTTGAGGGTGGGACCAGAAG	NA	ChIP-qPCR Smad/Oct Binding
Dppa5a-Smad3-F1	ATGCGCTTAGTGTGGAGGAC	NA	ChIP-qPCR Smad/Oct Binding
Lin28-Smad3-F1	CACCCCCTCCACATACACAC	NA	ChIP-qPCR

			Smad/Oct Binding
Klf4-Smad3-F1	CCTCACCCCTTTCTGTTCCC	NA	ChIP-qPCR Smad/Oct Binding
Sox2-Smad3-R1	ATGTGTGAGCAAGAACTGTCG	NA	ChIP-qPCR Smad/Oct Binding
Nanog-Smad3-R1	TTTTAACCACGGCTGCACCT	NA	ChIP-qPCR Smad/Oct Binding
Pou5f1-Smad3-R1	AGGAAGGGCTAGGACGAGAG	NA	ChIP-qPCR Smad/Oct Binding
Klf2-Smad3-R1	AGTCTAGCAGTGCCAAGCAG	NA	ChIP-qPCR Smad/Oct Binding
Sall1-Smad3-R1	GCGTCCTTATGAGCCCTCTC	NA	ChIP-qPCR Smad/Oct Binding
Sall4-Smad3-R1	TGCCTCTCCCTCCAACTTCT	NA	ChIP-qPCR Smad/Oct Binding
Fbxo15-Smad3-R1	TGCTTAGCAGCCAGGGATTC	NA	ChIP-qPCR Smad/Oct Binding
Gdf3-Smad3-R1	GCATTGGCATTCAGATGCGT	NA	ChIP-qPCR Smad/Oct Binding

Rex1-Smad3-R1	AGGCGATCGGGATTCTGAAG	NA	ChIP-qPCR Smad/Oct Binding
Jarid2-Smad3-R1	GAGACTGGAGGCCACTTGAC	NA	ChIP-qPCR Smad/Oct Binding
Dppa2-Smad3-R1	TCCTCATTCCCACATGGGTC	NA	ChIP-qPCR Smad/Oct Binding
Esrrb-Smad3-R1	TCAAAGCCACTCTGCTGGAG	NA	ChIP-qPCR Smad/Oct Binding
Dppa4-Smad3-R1	GACCTCAGCCGGCTTTTCAA	NA	ChIP-qPCR Smad/Oct Binding
Dppa5a-Smad3-R1	CCAAGGATGGCTGAGATGGT	NA	ChIP-qPCR Smad/Oct Binding
Lin28-Smad3-R1	CTTTGGATGCCCGCATTGAG	NA	ChIP-qPCR Smad/Oct Binding
Klf4-Smad3-R1	GTTGACTTTGGGGCTCAGGT	NA	ChIP-qPCR Smad/Oct Binding
Nanog-F1	TTCAGTCCCCGAAGAACCCA	NA	ChIP-qPCR H3K4me3
Dppa4-F1	GCTGGTCCCGGGATAAAACA	NA	ChIP-qPCR H3K4me3
Dppa5a-F1	TTTGGGTAGCCTCGGAAGGA	NA	ChIP-qPCR

			H3K4me3
Dppa2-F1	GGCCTCAATCTTGGAACCCT	NA	ChIP-qPCR H3K4me3
Sox2-F1	TTTGTCCGAGACCGAGAAGC	NA	ChIP-qPCR H3K4me3
Esrrb-F1	CCTCTGGCTACCACTACGGA	NA	ChIP-qPCR H3K4me3
Klf2-F1	AAGAGCTCGCACCTAAAGGC	NA	ChIP-qPCR H3K4me3
Utf1-F1	TCCCGGTGACTACGTCTGAT	NA	ChIP-qPCR H3K4me3
Lin28-F1	GCGCACGTTGAACCACTTA	NA	ChIP-qPCR H3K4me3
Lin28b-F1	TTACCTCCGCGGCCTCAT	NA	ChIP-qPCR H3K4me3
Jarid2-F1	CACATCCTTTGGCTTGCAGT	NA	ChIP-qPCR H3K4me3
Tcea3-F1	AGGATCGCCAAAAAGCTGGA	NA	ChIP-qPCR H3K4me3
Nanog-R1	GAGTGATAAGGACACCCGCTT	NA	ChIP-qPCR H3K4me3
Dppa4-R1	CCTCGGGTCCTCTCAGGTTA	NA	ChIP-qPCR H3K4me3
Dppa5a-R1	CCTGTAGCAGGCACTTTTGG	NA	ChIP-qPCR H3K4me3
Dppa2-R1	GCCTTCCTAGTCTGAGACGATT	NA	ChIP-qPCR H3K4me3
Sox2-R1	CTCCGGGAAGCGTGTACTTA	NA	ChIP-qPCR H3K4me3

Esrrb-R1	CCCCACCCAAAGCCTGAATC	NA	ChIP-qPCR H3K4me3
Klf2-R1	TCCATTTCCCTTCTGGAGGATG	NA	ChIP-qPCR H3K4me3
Utf1-R1	GAAGTAGCTCCGTCTCTCGG	NA	ChIP-qPCR H3K4me3
Lin28-R1	CCTCCCTATCTCCAGGTGGC	NA	ChIP-qPCR H3K4me3
Lin28b-R1	GAGCTCCGTCAGGAAGTGAC	NA	ChIP-qPCR H3K4me3
Jarid2-R1	AGAGCTTTATCGGACGCTGT	NA	ChIP-qPCR H3K4me3
Tcea3-R1	CATAGCTGGGGGAGTCTTGC	NA	ChIP-qPCR H3K4me3
Pou5f1-F1	CGAACCTGGCTAAGCTTCCA	NA	ChIP-qPCR H3K4me3
Sall1-F1	AGAGTCAGGCACAAAGTCCG	NA	ChIP-qPCR H3K4me3
Sall4-F1	GCCCTACATGTGACAAAGGC	NA	ChIP-qPCR H3K4me3
Fbxo15-F1	CCGATGGGCTGTGATCATTT	NA	ChIP-qPCR H3K4me3
Fgf4-F1	GACTACCTGCTGGGCCTCAA	NA	ChIP-qPCR H3K4me3
Pou5f1-R1	CAGTCCAACCTGAGGTCCAC	NA	ChIP-qPCR H3K4me3
Sall1-R1	CAGAGTTTCCGAGCTCCCAG	NA	ChIP-qPCR H3K4me3
Sall4-R1	ATTCGGATAGTGGCCGCAAG	NA	ChIP-qPCR

			H3K4me3
Fbxo15-R1	ACCTCATGGAACATCCATCCC	NA	ChIP-qPCR H3K4me3
Fgf4-R1	CCTGCAGGTGGAATCCGATG	NA	ChIP-qPCR H3K4me3
Gdf3-F1	AAGATCCAAGGCCAGACTCC	NA	ChIP-qPCR H3K4me3
Dnmt3l-F1	TTCCCGGGAGACACCTTCTT	NA	ChIP-qPCR H3K4me3
Nodal-F1	ATTGTTTCTCCGTGGGCAGG	NA	ChIP-qPCR H3K4me3
Cbx7-F1	GAGCAGCCTCACCTTCCG	NA	ChIP-qPCR H3K4me3
Cbx1-F1	ACCGATTCCTCTCGGACTCT	NA	ChIP-qPCR H3K4me3
Eras-F1	GTGTAAAGCTCGGGGTTGGA	NA	ChIP-qPCR H3K4me3
Tbx3-F1	GAGGCCAAGGAACTTTGGGA	NA	ChIP-qPCR H3K4me3
Prdm14-F1	GCATGCGCGTAGTAGGTAGT	NA	ChIP-qPCR H3K4me3
Rex1 (Zfp42)-F1	CTGCACTGCACACTCACTCT	NA	ChIP-qPCR H3K4me3
Tcf15-F1	AGTGTCTCCTGCCCTAGGAT	NA	ChIP-qPCR H3K4me3
Zfp296-F1	AAGCGTCAACTCCAAACTGC	NA	ChIP-qPCR H3K4me3
Lefty2-F1	CTGCCCTCATCGACTCTAGG	NA	ChIP-qPCR H3K4me3

Gdf3-R1	GTTCGTGGGAACCTGCTTCA	NA	ChIP-qPCR H3K4me3
Dnmt3l-R1	CCTGAGGATCTCACAGGATTTCA	NA	ChIP-qPCR H3K4me3
Nodal-R1	CCCACGCCCTACCCTTTTTA	NA	ChIP-qPCR H3K4me3
Cbx7-R1	GCATGGAGCTGTCAGCCATA	NA	ChIP-qPCR H3K4me3
Cbx1-R1	ATTCCTCGGTCTCCGCTCTT	NA	ChIP-qPCR H3K4me3
Eras-R1	TAGGGCCACCGGTCTAACTT	NA	ChIP-qPCR H3K4me3
Tbx3-R1	GAGGGGTCCCCCACTGATTA	NA	ChIP-qPCR H3K4me3
Prdm14-R1	CAGCTCGGTTTCCCACAGAT	NA	ChIP-qPCR H3K4me3
Rex1 (Zfp42)-R1	TTTCTGAATCAGGTCTCAACCA	NA	ChIP-qPCR H3K4me3
Tcf15-R1	TCTGGGCTCCCTCTTTCCTT	NA	ChIP-qPCR H3K4me3
Zfp296-R1	CCGAAGGTCCCGATACTAGC	NA	ChIP-qPCR H3K4me3
Lefty2-R1	CTATCATCTGCTGCTCCCCTC	NA	ChIP-qPCR H3K4me3

Table 2.4. qPCR primers

A table of primers used for various qPCR experiments, including ChIP-qPCR.

2.11 General Molecular Biology

2.11.1 SeraMag bead isolation of DNA from solution

We produce our own SeraMag bead solution, making a product similar to the commercially available Agencourt Ampure XP Bead cocktail. The general principle of the procedure is based on paramagnetic magnetite beads that are coated in negatively charged carboxyl molecules. Upon addition of the hydrophilic crowding molecule PEG (polyethylene glycol) along with salt (Na+) to a solution of DNA, the negatively charged DNA, shielded by Na+ ions, aggregates and precipitates out of solution, and also binds to the carboxyl groups of the beads. When introduced to a magnetic field, the beads are drawn out of solution, along with the DNA. Upon removal of PEG, and addition of H2O, the DNA is then re-hydrated and goes back into solution. The concentration of PEG can be adjusted by adding different volumes of 18% PEG to the initial solution of DNA, creating different stringency criteria for DNA-bead interaction, allowing for size selection of DNA fragments (Rohland et al, 2012). For example, adding the 18% PEG solution to a DNA containing solution at 1.2(PEG-bead solution):1(DNA solution) selectively excludes almost all DNA fragments less than 100bp in size, and a further reduced ratio of 0.9(PEG-beads):1(DNA) results in exclusion of any DNA fragments less than 200bp (Rohland et al, 2012). Thus, the lower the concentration of PEG and Na+, the larger the size of DNA fragment precipitated out of the solution.

The final Sera-Mag solution is as follows: 0.1% Sera-Mag Magnetic Speed-beads (FisherSci, cat.#: 09-981-123), 18% PEG-800 (w/v) (Sigma Aldrich cat.#: 89510), 1M NaCl, 10mM Tris-HCl ph8.0, 1mM EDTA, 0.05% Tween 20. The beads are stored at 4^{o}C, and there efficiency was tested using the Fermentas ladder (Fisher #FERSM1211).

2.11.2 Gibson assembly cloning protocol

Many cloning reactions were performed using the Gibson assembly protocol. The method was developed to assemble multiple fragments of DNA into a single linear strand in a 1-step reaction (Gibson et al, 2009). DNA fragments with, as small as 15bp,

overlapping sequences can be ligated together in the presence of T5 exonuclease (Epicentre), Phusion DNA polymerase (New England Biolabs) and Taq DNA ligase (New England Biolabs). The 5' T5 exonuclease chews, or removes, nucleotides from the 5' strand of fragments of DNA, creating a single stranded 3' overhang. When other DNA fragments are present, with identical end sequence, they will hybridize 3' overhangs. The Phusion polymerase can then fill the gaps in the annealed strands, and the Taq Ligase seals the nicks.

Gibson assembly reactions are initiated by adding equimolar amounts of PCR products and the desired plasmid backbone, which has been linearized with a restriction enzyme at the site of fragment insertion. Some reactions require the insert fragments to be at 3X to 10X the molarity of the plasmid backbone; a parameter that needs to be tested experimentally. The DNA mix is added to 10ul Gibson reaction mix (100mM Tris-HCl, 10mM MgCl2, 0.2mM dNTP, 0.5U Phusion Polymerase, 0.16U T5 exonuclease, stored at -20) and topped up to 20µl. The reaction is incubated at 50^{o}C for 1hr, and then placed on ice. 3ul of the reaction are transformed into competent bacteria.

2.11.3 Digestion of plasmid DNA with restriction enzymes

Restriction enzymes were used to 'cut and paste' various segments of DNA between plasmids, as well as to check correct orientation of inserts in newly constructed plasmids. A typical restriction enzyme reaction consisted of 100-1000 ng DNA, 1ul restriction enzyme (New England Biolabs), 1ul appropriate restriction enzyme buffer, topped up to 10ul with ddH_2O. The reactions are incubated at 37^{o}C for 1 hour unless otherwise specified by New England Biolabs. After digestion, the products were run on an agarose gel containing SybrSafe (Life technologies) to visualize the bands and/or cut out specific bands for cloning purposes. DNA fragment were isolated from agarose gels using the Zymoclean gel DNA recovery kit according to manufacturers instructions (Zymo Research).

2.11.4 Ligation of DNA fragments

Both blunt and sticky end DNA fragment ligations were accomplished by adding the plasmid backbone to insert at a concentration ratio of 1:3. A typical ligation reaction involved 50ηg Vector backbone, 150ηg insert, 1ul T4 DNA ligase (New England Biolabs), 1ul T4 Buffer, and then topped up to 10ul with ddH_2O. The reaction was either stored at 16^oC overnight, or incubated at room temperature for 1-hour prior to bacterial transformation.

2.11.5 Dephosphorylation of plasmid DNA

In ligation reactions, it is often necessary to dephosphorylate the plasmid backbone, to avoid backbone self-ligation. To accomplish this, plasmids were first digested, and then 1ul calf intestinal phosphatase (New England Biolabs) was added directly to the reaction for 30 minutes at 37^oC, prior to purification for ligation reaction.

2.11.6 Phosphorylation of DNA fragment

To aid certain ligation reactions, it is beneficial to have phosphorylated DNA inserts and dephosphorylated plasmid backbone. To phosphorylate the insert, ~1μg DNA fragments or 100μM oligos were added to 1μl T4 Polynucleotide Kinase (New England Biolabs), with 1ul T4 PNK Buffer, 1μl 32P ATP (New England Biolabs) topped up to 20ul with ddH_2O. The reaction was incubated at 37^oC for 30 minutes prior to gel extraction or cleanup with MinElute reaction cleanup kit (Qiagen) according to manufacturers protocol.

2.11.7 Preparation of cDNA

RNA was isolated from ~10^6 cells of 129 MEFs or TNG-MKOS ES or human neonatal dermal fibroblasts (HDFn, Life technologies) lines with TRIzol (Life Technologies) according to manufacturers instructions. The RNA was then reverse transcribed to

cDNA with the Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) (Life Technologies). 1µg of RNA was topped up to 5ul in ddH_2O and then added to a pre-reaction cocktail of 100uM Oligo dT, 100mM dNTPs to final volume of 10ul, for pre-incubation at 65°C for 5 minutes followed by incubation at 37°C indefinitely. The pre-incubated sample was then added to 5X M-MLV RT enzyme buffer, 2ul DTT, 1ul RNase Inhibitor (Life Technologies), 1ul M-MLV RT and topped up to 20ul in ddH2O for incubation at 37oC for 1 hour, followed by 10 minutes at 90°C and then 4°C indefinitely. The final product was diluted 10X in ddH_2O and then 2ul of that was used for PCR reactions.

2.11.8 Bacterial transformation

Roughly 10ηg of plasmid DNA was added to 50ul of chemically competent *E.coli* (Commercially available Top10 or Stbl3 lines from Life Technologies, or homemade DH5α or Stbl3 competent bacteria) and incubated on ice for 5 minutes, prior to incubation at 45°C for 1 minute, followed by 3 minutes on ice. Transformed bacteria were then added to 300ul of LB broth, and incubated at 37°C prior to plating on antibiotic containing LB agar plates for growth at either 37°C or 30°C overnight. The next day colonies were picked with a pipette tip (1000ul Tip) and placed into LB broth for bacterial expansion and plasmid DNA extraction. Plasmids were prepped from 5ml (mini-prep) or 100-250ml (midi-prep) bacterial broths that were grown in LB broth with antibiotic in a shaking incubator at 37°C or 30°C (recombination prone plasmids grown at 30°C) and isolated using QIAprep kits (Qiagen) according to manufacturers protocol.

2.11.9 Sequencing of plasmids

Plasmids were sequenced by Sanger sequencing with Edinburgh Genomics. Each sequencing reaction consisted of 200ng of plasmid, 1.6pmol primer, topped up to a 6ul reaction with ddH_2O. Primers for sequencing were often the same ones used for cloning, but in some case additional primers were needed, as indicated below.

Primers for sequencing	
Name	**sequence**
Smad3F400	CAGAGAGTAGAGACGCCAGT
Smad3R400	CGTGGCACCAACACTGGAGG
SmadF800	CACTGACCCCTCCAACTCGG
SmadR800	ACAGTAGGCCCAGGCAGAAG

Table 2.5. Primers for plasmids sequencing

A list of primers used to sequence plasmids

3: TGFβ signalling inhibition boosts early-intermediate reprogramming populations

3.1 Introduction

Soon after the discovery of iPSCs, a large body of work was dedicated to screening for small molecules that could boost the extraordinarily inefficient reprogramming process. In 2009, 4 independent groups identified that inhibition of TGF-β signalling resulted in 20 to 60-fold increase in reprogramming efficiency (Ichida et al., 2009a; Li et al., 2010a; Lin et al., 2009; Maherali and Hochedlinger, 2009). Whilst the mechanism of reprogramming enhancement by TGF-β inhibition was not clearly demonstrated, because of TGF-β's known roles in development, it was proposed to be in part due to an enhanced mesenchymal to epithelial transition (MET) (Lin et al., 2009; Maherali and Hochedlinger, 2009). However, several lines of evidence suggested that TGF-β inhibitors might contribute to reprogramming enhancement through pathways independent of the MET. For example, pre-treatment of fibroblasts with TGF-β inhibitor, as a strategy to alleviate mesenchymal signalling, does not have any effect on reprogramming, suggesting that the inhibitor acts in cooperation with the 4 reprogramming factors (Ichida et al., 2009a; Maherali and Hochedlinger, 2009). Secondly, when reprogramming epithelial cells, treatment with TGF-β inhibitors results in over 6-fold increase in reprogramming efficiency, again suggesting mechanisms independent of the MET (Li et al., 2010a). Finally, treatment of partially reprogrammed iPSCs, which have already undergone MET, with a short pulse of TGF-β inhibitor results in a marked increase in *Nanog* expression within 10-hours of treatment, driving the cells to a stable pluripotent state (Ichida et al., 2009a). It is clear that TGF-β inhibition has profound effects on the reprogramming process and that the mechanisms of action are, at least in part, independent of the MET.

3.1.1 Aims of this chapter

The ultimate aim of this chapter is to assess which reprogramming stages are most affected by the TGF-β inhibitor. It is possible that TGF-β inhibition might have multiple affects at various stages of reprogramming. To assess the impact of TGF-β inhibition during reprogramming, a time-course analysis is performed using FACS analysis coupled with surface markers and a pluripotency reporter, to observe reprogramming progression in the presence or absence of TGF-β inhibitor. Immunofluorescence microscopy and qPCR are used to investigate phenotypes in further detail at the various stages of reprogramming. Finally, populations of interest are FACS sorted to assess the affects of the TGF-β inhibitor on reprogramming sub-populations.

3.2 Results

3.2.1 TGF-β inhibition boosts reprogramming efficiency

The reprogrammable MEFs used for experiments throughout this chapter were produced as previously published (Figure 3.1) (Chantzoura E.; Skylaki, in press). Briefly, an Sp3 locus targeting vector containing the 4 reprogramming factors c-Myc, Klf4, Oct3/4, and Sox2 (MKOS) under a doxycycline (dox) inducible promoter followed by ires-mOrange and a CAG-rtTA expression cassette, was transfected into a *Nanog*-GFP mES reporter cell line (Chantzoura E.; Skylaki, in press). The resultant TNG-MKOS line was then used to produce chimeric embryos via morula aggregation, followed by MEF extraction at embryonic stage 12.5 (E12.5). The extracted MEF are a mixed batch of transgenic and wt cells, as each embryo has different levels of transgenic cell contribution. The percent of transgenic cells was assessed for each batch of MEF by monitoring mOrange expression via FACS analysis 48 hours after dox administration before freezing down. For reprogramming experiments, the proportion of transgenic MEF was adjusted to 5%, surrounded by 95% wt MEF; the previously established

optimal reprogramming conditions. Upon dox induction of the 4 reprogramming factors, the cells reprogram to an iPSC state within 12 days, as assessed by *Nanog*-GFP^{+} colonies, at an efficiency of ~1%.

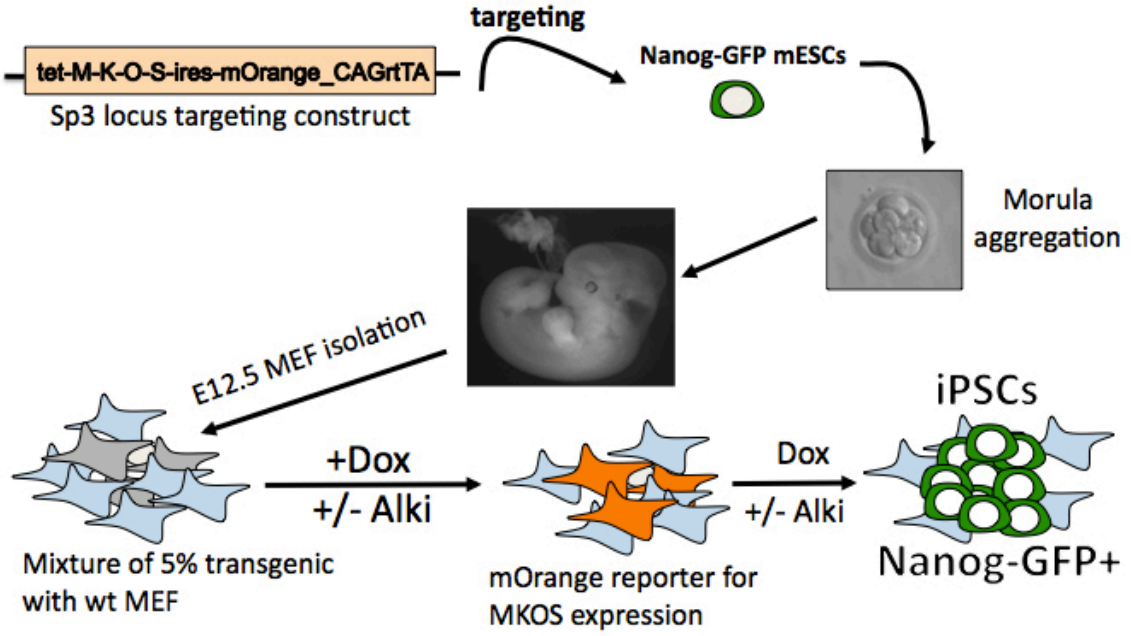

Figure 3.1. Transgenic reprogramming system

Our reprogramming system utilizes an ES line with a *Nanog*-GFP reporter, targeted with a dox inducible MKOS (cMyc, Klf4, Oct4, Sox2) reprogramming cassette and an IRES mOrange reporter. The ES line was then used for morula aggregation after which embryos were harvested at E12.5 and fibroblasts (MEF) were isolated. Transgenic fibroblasts are then mixed with wt fibroblasts to have 5% starting transgenic population at the outset of reprogramming. Upon addition of dox, the transgenes MKOS are expressed along with the mOrange reporter, allowing the tracking of cells as they reprogram to pluripotency. Resultant iPSCs can be distinguished based on morphology and expression of the Nanog-GFP reporter.

Previous reports have demonstrated that addition of Vitamin C (VitC) can vastly enhance reprogramming efficiency (Chen et al., 2013a; Esteban et al., 2010; Gao et al., 2015). The mechanism of VitC action during reprogramming was demonstrated to be in part due to the up-regulation of the micro RNA (miRNA) cluster 302-367, which is thought to enhance the MET (Gao et al., 2015). Because TGF-β inhibition was also proposed to promote MET, the TGF-β receptor inhibitor Alki (A83-01) was used to test if TGF-β inhibition could enhance reprogramming in the presence of VitC. Alki is an inhibitor of TGF-β Type I cell surface receptors, the receptors necessary for downstream TGF-β signal transduction (Tojo et al., 2005).

The addition of VitC results in over 10-fold increase in total colony numbers, as assessed by microscopy, counting colonies based on morphology and the mOrange reporter of transgene expression (Figure 3.2A). However, only 65% of colonies generated in the presence of VitC activate the endogenous *Nanog* promoter, as indicated by counting reporter *Nanog*-GFP^{+} colonies at day 12 of reprogramming. Addition of the Alki alone results in a 7-fold increase in total colony numbers, with nearly 100% of the colonies expressing *Nanog*-GFP (Figure 3.2A). Addition of both Alki and VitC results in a cumulative affect, whereby a 10-fold increase in reprogramming efficiency is observed and ~90% of colonies are *Nanog*-GFP^{+} at day 12 (Figure 3.2A). Because Alki was able to boost reprogramming efficiency in the presence of VitC, the remainder of reprogramming experiments were performed with VitC to have the most efficient reprogramming system and to identify novel mechanisms specific to TGF-β pathway inhibition.

Observations of colonies reprogramming in the presence of Alki indicated that the transgenic cells expanded rapidly in the presence of the inhibitor. To quantify this phenotype, FACS analysis was performed to monitor the proportion of mOrange^{+} (transgenic) cells, as compared to mOrange^{-} (wt) cells, in culture throughout reprogramming. Absolute value of transgenic and wt cells was extrapolated based on total cell counts from each well. In the presence of Alki, transgenic cells rapidly proliferated and were twice that of control conditions by day 2 of reprogramming and continued to expand more rapidly than the control at day 4 and 6 (Figure 3.2b).

Throughout the entire reprogramming process, there were generally less wt cells in the dish in the presence of Alki, indicating that the Alki acts in combination with the reprogramming factors to boost transgenic cell proliferation.

Previous work has hypothesized that TGF-β inhibition boost MET during reprogramming. To assess acquisition of epithelial character, FACS analysis was performed to quantify E-CADHERIN (E-CAD) levels during reprogramming. In our reprogramming system, E-CAD+ cell populations reached over 70% by day 4 of reprogramming in either the presence or absence of Alki (Figure 3.2C and D). This result indicated that the MET is not a major barrier to reprogramming in this system, and also that Alki can increase reprogramming efficiency independent from the enhancement of MET.

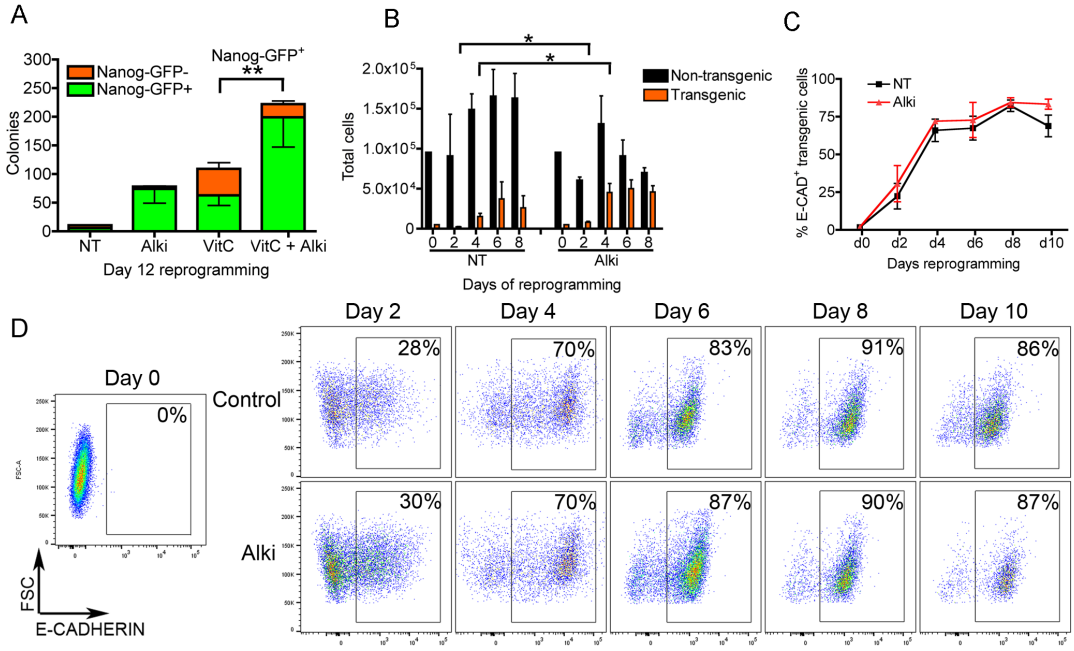

Figure 3.2. The Alki enhances reprogramming independent of the mesenchyme to epithelial transition

A) MEFs were reprogrammed with addition of Vitamin C (VitC) or Alki alone or in combination. Colonies were counted based on mOrange expression (all colonies) and *Nanog*-GFP+ (successfully reprogrammed) at day 12 of reprogramming. B) Total cells were counted and then FACS analyzed to extrapolate the numbers of transgenic cells (mOrange+) and wt cells (mOrange-) when reprogramming with or without Alki. Results display 3 independent experiments. C) Quantified FACS analysis results from 3 independent experiments analysing E-CADHERIN+ transgenic cells (mOrange+) during reprogramming. D) Example E-CADHERIN FACS analysis experiment, where only mOrange+ cells are displayed and quantified for E-CADHERIN expression. Error bars indicate standard deviation (s.d.) from 3 independent experiments. *P-value <0.05, **<0.01, based on a two-sided t-test.

One of the major roadblocks at the onset of reprogramming is the activation of senescence (Banito et al., 2009; Kawamura et al., 2009; Li et al., 2009). The rapid expansion of cells in the presence of Alki suggested that the TGF-β pathway might modulate senescence during the initial stages of reprogramming. Two major branches of senescence, $p19^{ARF}$-p53 and the p16 pathway, were investigated by confocal microscopy and qPCR during reprogramming. At day 4 of reprogramming, there is a marked increase in $P19^{ARF}$ protein in cells expressing the 4 factors ($mOrange^{+}$), but not in the presence of the Alki (Fig 3.3A and B). By day 8, levels of p19 within the colonies, both with and without the Alki, are at very low to undetectable levels (Figure 3.3A). The mRNA transcript level was assessed by qPCR, confirming that up-regulation of p19 expression at day 4 is bypassed in the presence of the Alki (Figure 3.3B). Both protein and transcript levels of p16 were very low to undetectable throughout reprogramming either in the presence or absence of Alki (Figure 3.3A and B). These results revealed that TGF-β signalling impeded the early stages of reprogramming by activating the $p19^{ARF}$ mediated senescence pathway.

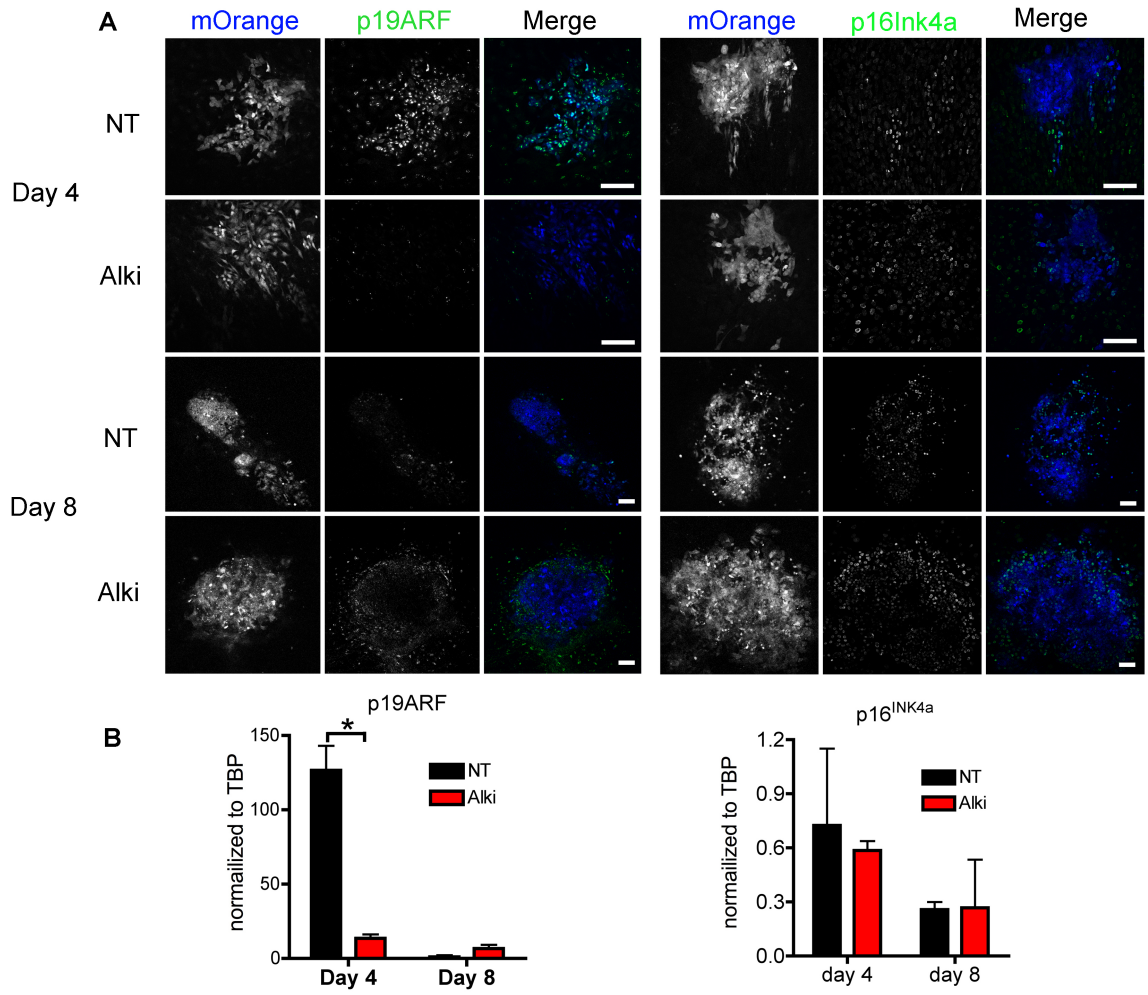

Figure 3.3. Reprogramming initiated p19^{ARF} mediated senescence is bypassed with addition of the Alki

A) Confocal microscopy imaging of single colonies (mOrange) stained for p19^{ARF} or p16^{Ink4a} at day 4 or 8 of reprogramming. Scale bars are 100μM. B) qPCR at day 4 and 8 of reprogramming with or without Alki, normalized to Tata Box binding Protein (TBP). Error bars indicate standard deviation (s.d.) from 2 independent experiments. *P-value <0.05 based on a two-sided t-test.

It is possible that multiple stages of reprogramming are affected by TGF-β inhibition. Indeed, while the presence of the inhibitor throughout the entire process has

the most profound effect (Figure 3.4), addition for only the first 6 days of reprogramming results in half of the effect of addition throughout the entire process. Whereas delayed addition of the Alki at day 6 of reprogramming results in over 4-fold increase in reprogramming efficiency (Figure 3.4). Those results indicate the inhibitor may have affects on multiple stages of reprogramming.

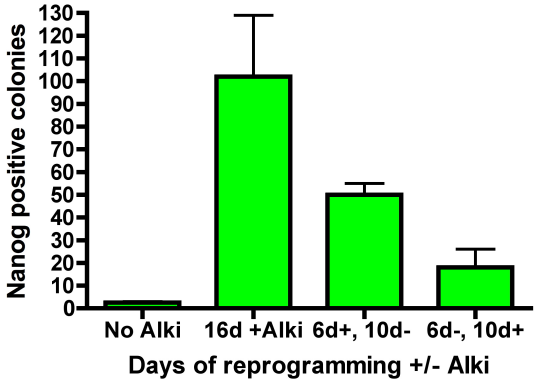

Figure 3.4. Affect of the Alki addition at various stages of reprogramming

Reprogramming was performed in presence of Alki for the entire experiment (16d+Alki), only the first 6 days (6d+, 10d-), or the last 10 days (6d-, 10d+). Nanog-GFP+ colonies were counted at day 16. Error bars indicate standard deviation (s.d.) from 2 independent experiments.

To investigate various stages of reprogramming, in addition to the *Nanog*-GFP reporter, the surface markers CD44 and Icam1 can be utilized to monitor reprogramming progression (O'Malley et al., 2013). As cells progress from fibroblast to iPSCs, *CD44* is gradually down regulated, followed by up-regulation of *Icam1* and *Nanog* as cells reach a pluripotent state (Figure 3.5). Another indication of successful reprogramming is the silencing of transgenes in iPSCs (Hotta and Ellis, 2008; Okita et al., 2007), which can be monitored by the mOrange reporter in this reprogramming system (Figure 3.5).

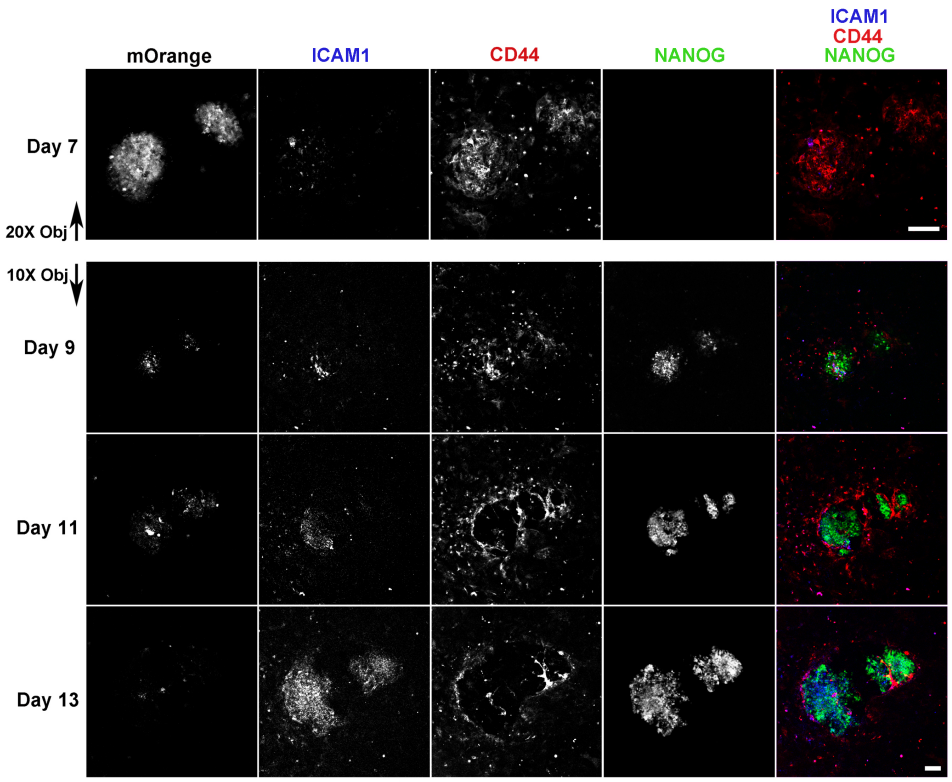

Figure 3.5. Live-cell imaging with reprogramming stage specific surface markers

Confocal microscopy live imaging of colonies (mOrange+) undergoing reprogramming. Fluorophore conjugated antibodies were used to visualize Icam1 and CD44, in conjunction with the *Nanog*-GFP reporter on a pair of colonies undergoing reprogramming. Scale bars are 100 μm.

Reprogramming progression was monitored by FACS analysis with CD44, ICAM1 and *Nanog*-GFP over a 16 day reprogramming experiment. In the presence of Alki, there is a clear acceleration of CD44 down regulation at day 8 of reprogramming. Furthermore, ICAM1 up regulation and *Nanog*-GFP activation is apparent by day 8 of reprogramming; 2 to 4 days before the control experiment (Figure 3.6). At the final time points of reprogramming, at day 12 and 16, there remain populations of cells and colonies that express mOrange but are unable to up regulate *Icam1* or *Nanog*, stuck at

intermediate stages of reprogramming. In the Alki condition, the stuck populations are noticeably fewer, indicating that roadblocks are overcome in the presence of the inhibitor (Figure 3.6).

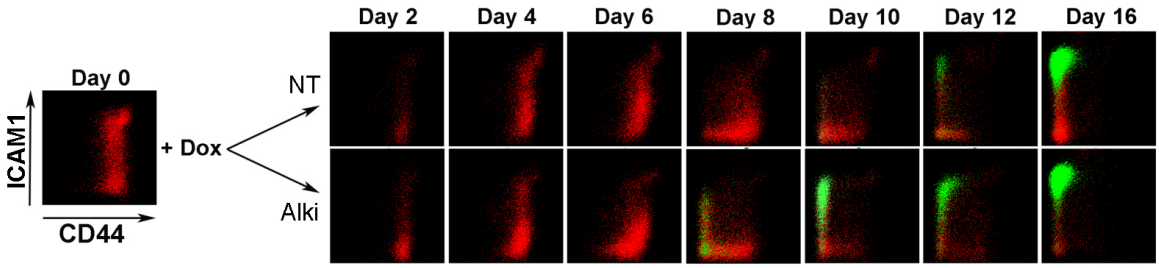

Figure 3.6. FACS analyses with CD44 and ICAM1 reprogramming with or without Alki.

A reprogramming time course FACS analyses of CD44 and ICAM1 staining, displaying all transgenic cells (mOrange+). *Nanog*-GFP positive cells are demarked by green dots, while all other transgenic cells are in red.

Because the reprogramming populations behave so differently in the presence of the Alki, it is hard to distinguish which populations of cells are most affected by the inhibitor. To identify reprogramming populations targeted by the inhibitor, cells were reprogrammed in the presence or absence of the Alki for 10 days and then sub-populations were sorted, based on CD44/ICAM1/Nanog-GFP expression, into media without Alki to assess there ability to form colonies after a further 10 days of reprogramming (Figure 3.7). Cells that were at intermediate stages of reprogramming, $CD44^{-}/ICAM1^{-}/NANOG\text{-}GFP^{+}$ (2N+) and $CD44^{-}/ICAM1^{+}/NANOG\text{-}GFP^{-}$ (3N-), showed over 2-fold increase in reprogramming efficiency when sorted from reprogramming in the presence of the Alki for the first 10 days. Interestingly, the early $CD44^{+}/ICAM1^{-}/$*Nanog*$\text{-}GFP^{-}$ populations coming from Alki conditions were less

efficient at reprogramming, perhaps as a result of sorting at such a late time-point of reprogramming (Figure 3.7).

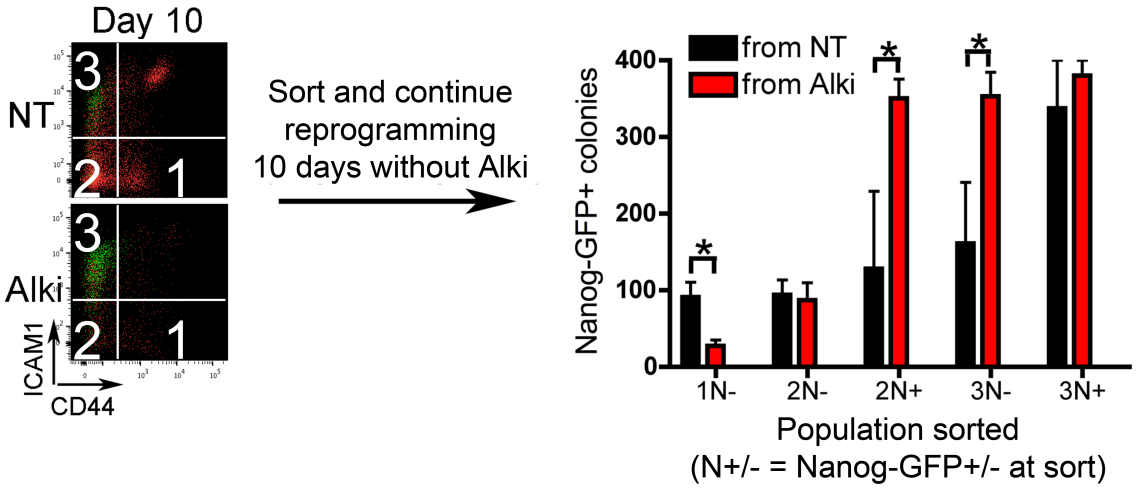

Figure 3.7. Alki affects intermediate reprogramming populations

Day 10 reprogramming populations were sorted based on CD44, ICAM1 and *Nanog*-GFP (N+/-) expression profiles. Population 1 (CD44+/ICAM1-), 2 (CD44-, ICAM1-), and 3 (CD44-/ICAM1+) represent reprogramming populations sorted. Each population was also sorted based on whether or not it expressed *Nanog*-GFP (N+ or N-). 10 days after sorting, *Nanog*-GFP+ colonies were counted and displayed in graph on the right. Error bars indicate standard deviation (s.d.) from 3 independent experiments. *P-value <0.05 based on a two-sided t-test.

3.3 Discussion

3.3.1 TGF-β signalling inhibition acts independent of the mesenchymal to epithelial transition during reprogramming

The early stages of reprogramming are characterized by a mesenchymal to epithelial transition (MET), as fibroblasts convert to a pluripotent state (Li et al., 2010a; Samavarchi-Tehrani et al., 2010). E-cad expression is a hallmark of the MET, and is necessary for reprogramming as a crucial component of the stem cell epithelial program (Redmer et al., 2011). Because the TGF-β signalling pathway is known to be involved in the specification of mesenchyme during development (Xu et al., 2009), it was proposed that TGF-β inhibition accelerates reprogramming by alleviating mesenchymal signalling and boosting the transition to epithelial character (Polo and Hochedlinger, 2010). However, results presented in figure 3.2 demonstrate that E-cad expression is not significantly affected by TGF-β inhibition in our reprogramming system. Furthermore, addition of the Alki at d6, post MET, or sorting cell populations beyond MET coming from Alki condition, demonstrated enhanced reprogramming efficiency. These data suggested that the major mechanism by which TGF-β inhibition boosts reprogramming is not the acquisition of epithelial character.

Treatment of epithelial cells in culture with TGF-β ligands, can convert them to a mesenchymal phenotype (Miettinen et al., 1994; Piek et al., 1999; Valcourt et al., 2005). During reprogramming, there are TGF-β ligands in the culture, sourced from both the fibroblasts themselves, and present in serum of the media (Chin et al., 2001; Li et al., 2010b; Schmid et al., 1998). Therefore Alki should have shown some sort of MET acceleration during reprogramming. However, probably the over-expression of *Klf4*, a master transcription factor of the epithelial program (Garrett-Sinha et al., 1996), overrides any effect of TGF-β ligands in culture, even in the absence of Alki. The finding that *E-cad* is expressed at maximal levels by day 4 of reprogramming also suggests that the MET is not a substantial barrier to overcome on route to pluripotency in our reprogramming system. Our lab previously demonstrated that most polycystronic reprogramming vectors commonly used in the research field contain a truncated form of

Klf4, missing the first 9 aa of full length Klf4 (Chantzoura E.; Skylaki, in press). With such polycystronic reprogramming vectors, MET seems to be a roadblock for reprogramming. However, with a polycystronic reprogramming vector with full length Klf4, MET is not a roadblock and thus, any affect of the Alki on MET might be subtle during such an efficient process.

3.3.2 TGF-β inhibition enables bypassing of senescence at onset of reprogramming

During reprogramming the over expression of the oncogene *c-Myc*, and to a lesser extent *Oct4* and *Sox2*, elicits a senescence response in MEF that includes the up regulation of $p19^{ARF}$, $p16^{Ink4a}$ and *p53* (Banito et al., 2009; Hong et al., 2009; Kawamura et al., 2009; Li et al., 2009). Senescence, by definition, is a cellular program characterized by the irreversible arrest of the cell cycle, and is therefore a major barrier to successful reprogramming (Kawamura et al., 2009; Li et al., 2009). In parallel, extensive studies have provided strong links between TGF-β and senescence pathways in various contexts (Cordenonsi et al., 2003; Senturk et al., 2010). Throughout reprogramming, cells would be exposed to TGF-β, as it is present in the serum of media and is produced by fibroblasts (Chin et al., 2001; Li et al., 2010b; Schmid et al., 1998). Interestingly, the $p19^{ARF}$ locus is a direct target of SMAD2/3 and is activated upon TGF-β stimulation of MEFs (Zheng et al., 2010). Treating MEFs with inhibitors of TGF-β receptors, as well as knockdown of Smad2/3, impedes TGF-β mediated $p19^{ARF}$ induction (Zheng et al., 2010). In addition, TGF-β receptors activate the p38MAPK/NF-κB pathway, which initiates a senescence response in fibroblasts (Freund et al., 2011). Blocking of TGF-β signalling during reprogramming possibly alleviates p38MAPK-initiated senescence, and certainly bypasses $p19^{ARF}$ activation, as demonstrated in this thesis.

When *c-Myc* alone is over-expressed in MEF, there is a marked increase in proliferation, and some MEF transform to an immortalized self-renewing state, termed

transformation (Greenberg et al., 1999; Qi et al., 2004b). $P19^{ARF}$ directly mediates C-MYC activity, binding to C-MYC and inhibiting its ability to activate transcription of target loci. Over-expression of $P19^{ARF}$ results in a failure of C-MYC to induce proliferation and transformation of MEF, in a process that is independent of p53 (Qi et al., 2004b). In the reprogramming context, it is plausible that $p19^{ARF}$ up regulation not only activates the senescence program, but also inhibits the exogenous C-MYC activity. It would be interesting to test if $p19^{ARF}$ knockdown, or Alki treatment, can enhance reprogramming in absence of *c-Myc* expression. Interestingly, when reprogramming with $p19^{ARF}$ and $p16^{Ink4a}$ knockdown, the increase in reprogramming efficiency was far greater in 4-factor than in 3-factor reprogramming system without *c-Myc* (Li et al., 2009). The same experiment was not performed with $p19^{ARF}$ knockdown alone, but those findings suggest there may be $P19^{ARF}$ direct inhibition of C-MYC activity during reprogramming.

3.4 The Alki accelerates transition from a partially reprogrammed to a pluripotent state

Reprogramming is a step-wise process, with cells gradually overcoming barriers, and establishing the pluripotency transcriptional network (Ruetz and Kaji, 2014). To assess reprogramming progression, cell surface markers and endogenous pluripotency reporters are used. One of the most stringent pluripotency markers is Nanog, which is activated relatively late in the reprogramming process, and strongly correlates with a stable pluripotent state (Buganim et al., 2012; O'Malley et al., 2013). In addition, the surface markers CD44 and Icam1 demarcate the various stages of reprogramming, allowing for precise monitoring of reprogramming progression (Lujan et al., 2015; O'Malley et al., 2013; Zunder et al., 2015). Fibroblasts express high levels of CD44 and a broad range of ICAM1 expression. As populations transition to a partially reprogrammed state, they down regulate *Icam1* and *CD44* expression, to a double negative intermediate, partially reprogrammed state. Eventually, the intermediate populations begin to activate Icam1 expression in conjunction with *Nanog* expression. The end-stage iPSC profile is that of low *CD44*, high *Icam1* and high *Nanog* expression

(O'Malley et al., 2013). With staining for CD44, ICAM1, and a *Nanog*-GFP endogenous reporter, I investigated the affect of Alki on various reprogramming populations. The results highlighted the inhibitor's role in not only boosting efficiency, but also acceleration of reprogramming process. By day 8 of reprogramming, the double negative CD44/ICAM1 intermediate stage is rapidly produced in the presence of inhibitor, and *Nanog*-GFP expression is observed 4 days earlier in the process.

In addition, the sorting experiments revealed that intermediate populations that came from Alki conditions were more likely to generate *Nanog*-GFP^{+} colonies even if they have the same CD44/ICAM1/*Nanog*-GFP marker expression, suggesting those populations had unique characteristics generated by the presence of the Alki. Those results demonstrate the affects of TGF-β inhibition at intermediate stages of reprogramming, beyond MET and senescence.

Previous work has highlighted a role for Alki in producing a more homogenous pluripotent state, whereby treatment of mouse ES cells results in a high *Nanog* expression (Galvin-Burgess et al., 2013). Treatment of ES cells with the Alki results in activation of the BMP pathway and associated response genes, such as Id proteins, to become up regulated and drive pluripotency gene expression (Galvin-Burgess et al., 2013). In the reprogramming context, a short pulse with TGF-β inhibitor results in a marked increase in NANOG protein levels within 12 hours of treatment, and drives partially reprogrammed cells to pluripotency (Ichida et al., 2009a). The elevation of BMP signalling in the presence of Alki could be a potential mechanism through which Alki might enhance reprogramming populations transitioning to a pluripotent state. Indeed, it has previously been demonstrated that addition of BMPs to culture can boost reprogramming efficiency (Samavarchi-Tehrani et al., 2010). Future work could explore if a BMP response is observed during reprogramming in the presence of TGF-β inhibition.

4: Identifying the underlying mechanism of TGF-β inhibition during reprogramming

4.1 Introduction

TGF-β inhibition has profound effects on reprogramming, increasing the overall efficiency over 10-fold and vastly accelerating the transition to a pluripotent state. TGF-β receptors target a number of downstream pathways including p38(MAPK), PI3K(Akt1/2), RhoA, Ras(ERK), TAK1 and Smad2/3 (Shi and Massague, 2003b; Zhang, 2009). Upon TGF-β ligands binding receptors, the receptors dimerize and the cytosolic kinase domains become active (Shi and Massague, 2003a) (Figure 4.1). The most well-established and extensively characterized target of TGF-β receptor activity is the phosphorylation and activation of Smad2/3 transcription factors (Figure 4.1). Indeed, Smad2/3 are core players in TGF-β downstream signalling, whereby knockout of either protein greatly reduces the TGF-β transcriptional response (Kretschmer et al., 2003; Lakos et al., 2004; Wakabayashi et al., 2011). This chapter tests the hypothesis that dampening Smad2/3 signalling is the primary downstream mechanism of action of the TGF-β receptor inhibitors in the reprogramming context.

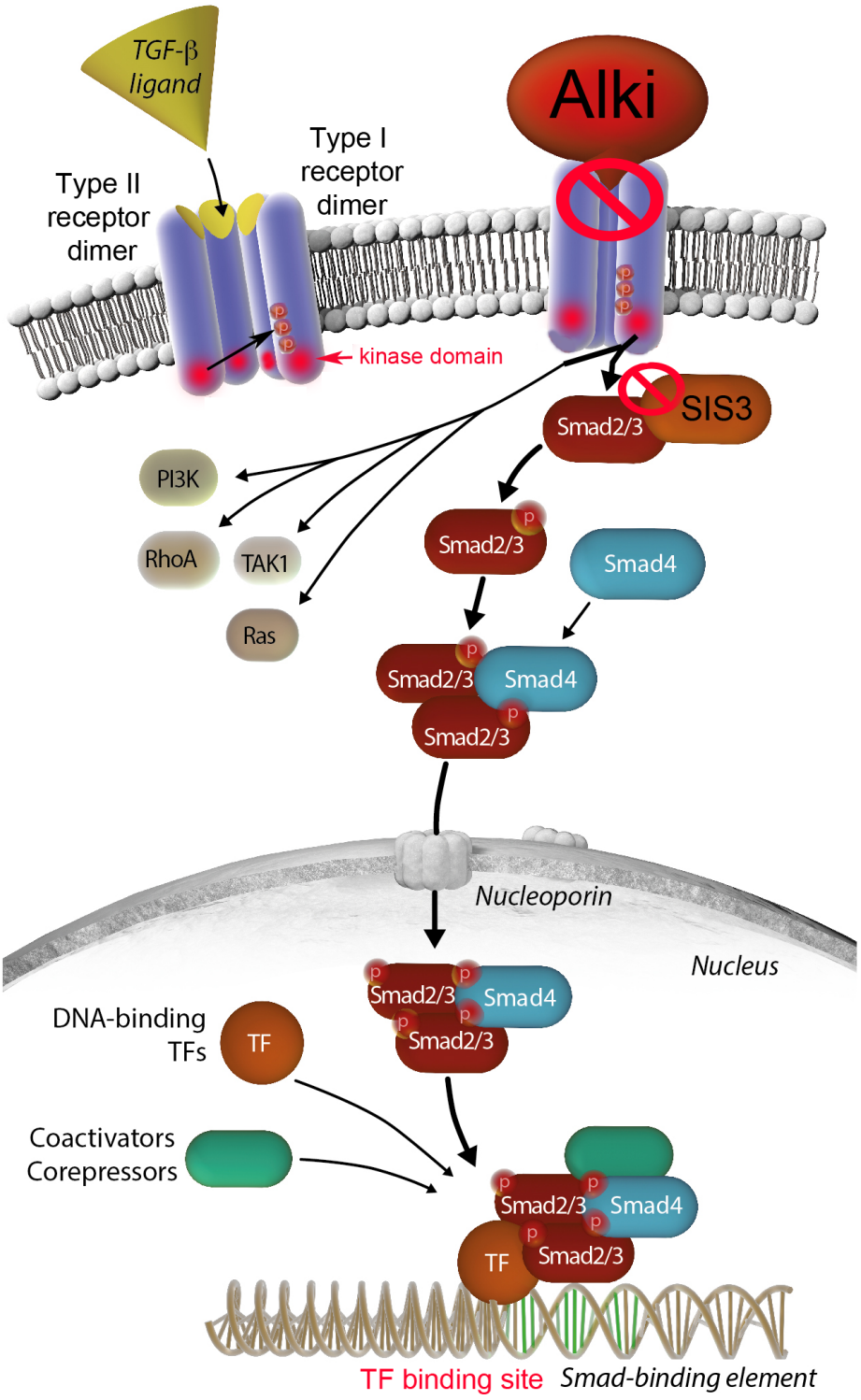
TGF-β ligand
Alki
Type I receptor dimer
Type II receptor dimer
kinase domain
SIS3
Smad2/3
PI3K
RhoA
TAK1
Ras
Smad2/3
Smad4
Smad2/3
Smad4
Smad2/3
Nucleoporin
Nucleus
Smad2/3
Smad4
Smad2/3
DNA-binding TFs
TF
Coactivators
Corepressors
Smad2/3
Smad4
TF
Smad2/3
TF binding site
Smad-binding element

Figure 4.1. Overview of TGF-β signalling and associated inhibitors

TGF-β signalling is initiated by ligands binding to Type II membrane receptor dimers, which then form a tetramer with and activate (phosphorylate) Type I receptor dimers. Type I receptors then phosphorylate and activate downstream signalling pathways, including PI3K, RhoA, Ras, Tak1 and Smad2/3. The Alki works by inhibiting the Type I receptor kinase domain, thereby inhibiting all Alk receptor downstream signalling. The specific inhibitor of Smad3 (SIS3) inhibits only the Smad3 signalling pathway.

4.1.1 Aims of this chapter

The ultimate aim of this chapter is to assess the role of Smad2/3 during reprogramming. A specific inhibitor of Smad3 (SIS3) is used to assess Smad3's importance for reprogramming progression. Smad3 inhibition phenotypes are explored with a combination of immunofluorescence, FACS analysis, western blotting and image cytometry. As a complementary approach, over-expression of constitutively active Smad3 (Smad3CA) and Smad2 (Smad2CA) are performed to further explore their roles during reprogramming. Finally, time course RNA sequencing was performed to assess transcription dynamics throughout reprogramming in the presence of Smad2CA/3CA.

4.2 Results

4.2.1 Inhibition of Smad3 impedes reprogramming

One of the most extensively characterized and primary targets of TGF-β receptors are the Smad2/3 transcription factors. To test if lack of the Smad2/3 mediated signalling is the main cause of the enhanced reprogramming phenotype by Alki, I added a specific inhibitor of Smad3 (SIS3) during reprogramming, and evaluated if it recapitulated the Alki results (Figures 4.1 and 4.2). Strikingly, and unexpectedly, the

presence of SIS3 during reprogramming almost completely abolished successful reprogramming (Figure 4.2). Furthermore, addition of SIS3 in the presence of Alki eliminates the positive affects of the Alki, again reducing reprogramming efficiency (Figure 4.2). The SIS3 did not have a general effect on fibroblast survival, as indicated by the presence of over 50 partially reprogrammed colonies at day 12 (Figure 4.2 and 4.3A), and counting of transgenic (mOrange^{+}) and wt (mOrange^{-}) cells at day 8 of reprogramming using FACS analysis (Figure 4.3B). There were similar levels of transgenic cells in control and SIS3 treatment at day 8 of reprogramming, while there were almost 2-fold more wt cells in the presence of SIS3, indicating the inhibitor did not affect fibroblast survival and might be beneficial to proliferation (Figure 4.3B).

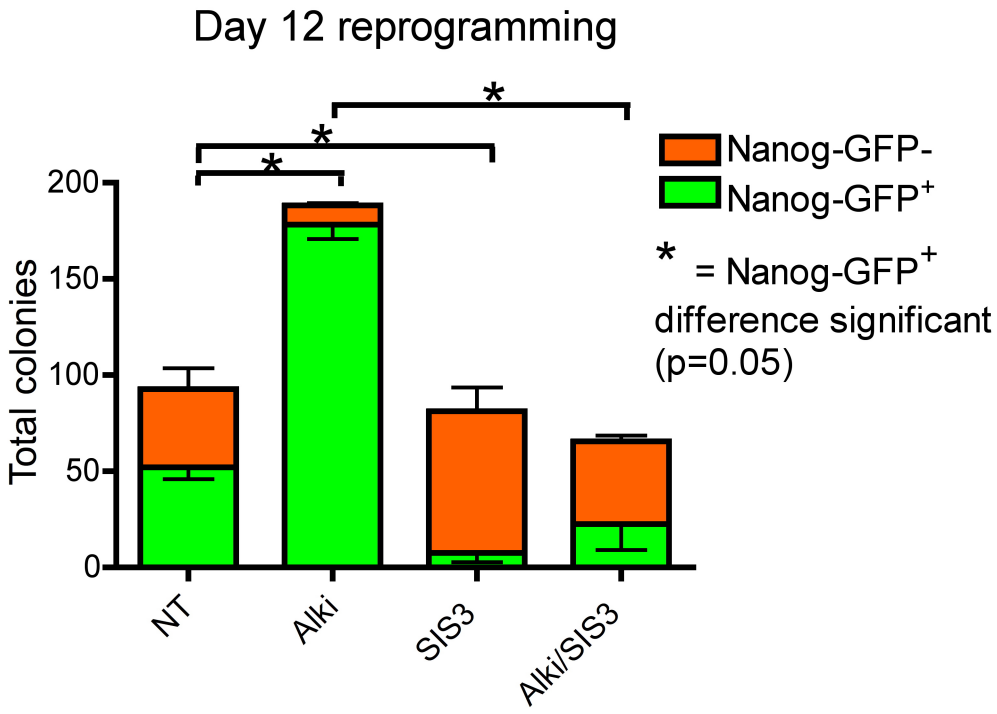

Figure 4.2. Specific inhibitor of Smad3 blocks reprogramming

Reprogramming was performed with addition of Alki, or a specific inhibitor of Smad3 (SIS3), alone or in combination. Colonies were counted at day 12 of reprogramming. Error bars indicate standard deviation (s.d.) from 3 independent experiments with 2 experimental replicates. *P-value <0.05 based on a two-sided t-test.

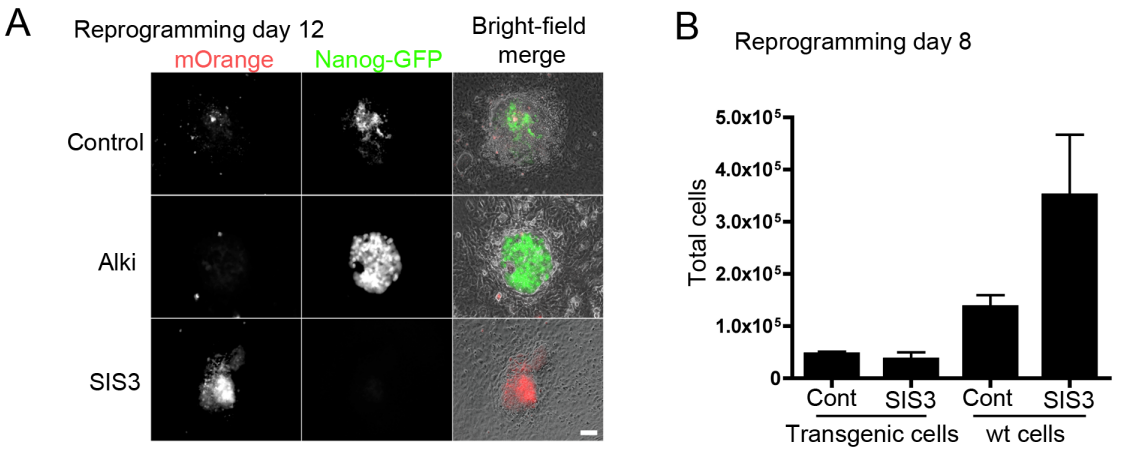

Figure 4.3. SIS3 blocks reprogramming independent of cell viability

A) Example colonies at day 12 of reprogramming. Scale bar 100μM. B) Cell counts based on FACS analysis from day 8 of reprogramming. Total cell were counted from 1 well of a 6-well reprogramming experiment, then FACS analyzed to enumerate transgenic (mOrange+) and wt (mOrange-) cells in the control (cont) and SIS3 treatments. Results show 2 independent experiments, with 2 experimental replicates. Error bars represent standard deviation.

To assess at what stage of reprogramming cells failed in the presence of SIS3, FACS analysis with CD44 and ICAM1 was performed from day 6-14 of reprogramming. By day 10 of reprogramming, a clear difference in populations emerged, as SIS3 reprogramming experiments stalled at the intermediate (CD44-/ICAM1-) stage of reprogramming, while the control conditions had started to up regulate ICAM1 and the Alki condition were also starting to express *Nanog*-GFP (Figure 4.4). By day 12 only

12% of cells made the final stage in the presence of SIS3, as compared to 53% and 68% of control and Alki conditions respectively (Figure 4.4).

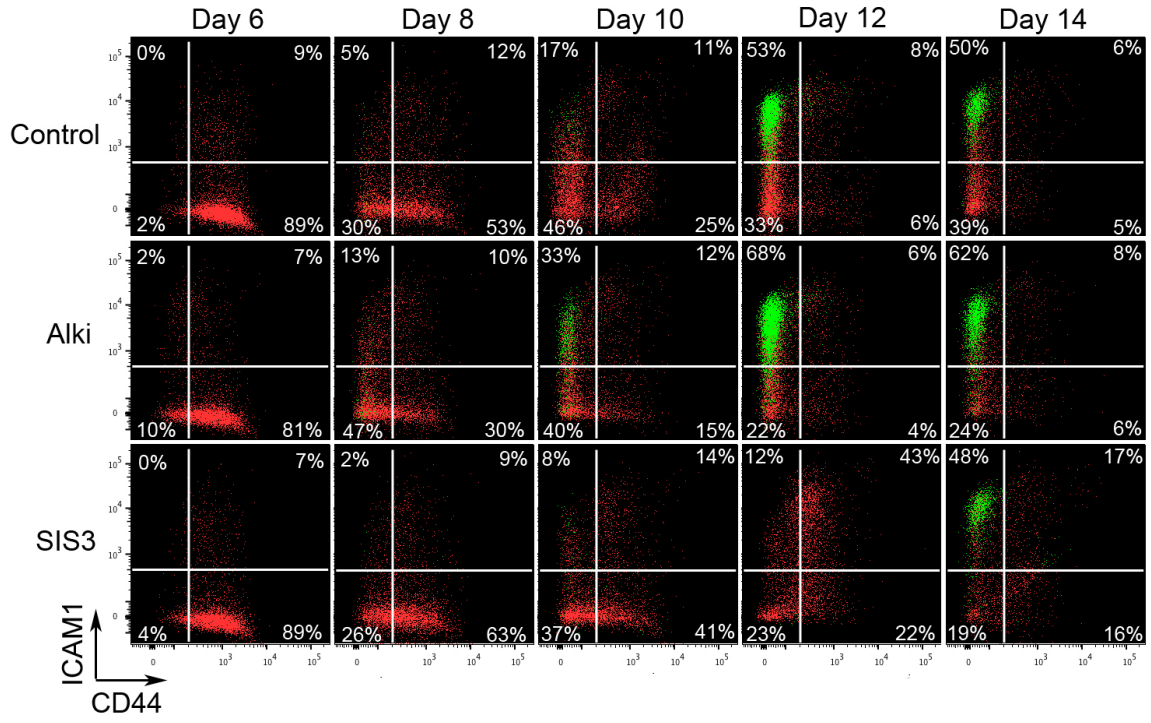

Figure 4.4. SIS3 stalls reprogramming at intermediate stage

FACS analysis of transgenic cells stained with CD44 and ICAM1 surface markers performed over a reprogramming experiment. Green dots represent *Nanog*-GFP+ cells, with red dots representing the rest of the transgenic (mOrange+) cells.

The SIS3 treatment results suggested that active Smad3 might be important for reprogramming, but this idea contradicted the effect of Alki, which was supposed to block SMAD2/3 phosphorylation. To confirm the amount of activated Smad3 during reprogramming in the presence and absence of the inhibitors, immunofluorenece was performed with a c-terminal phosphorylated-SMAD3 (pSMAD3) specific antibody. At day 4 and 8 of reprogramming, low levels of pSMAD3 were detected in the untreated

reprogramming context (Figure 4.5 and 4.6). The colonies in +SIS3 culture had low to undetectable levels of pSmad3, as expected. By day 12 of reprogramming, pSmad3 levels are at low to undetectable levels (Figure 4.7). In contrast, surprisingly, there was an apparent increase of pSmad3 in the presence of the Alki (Figure 4.5 and 4.6). Interestingly, pSmad3-high cells co-localized in regions of the colony with higher levels of Nanog-GFP expression in the absence of the inhibitors and presence of Alki (Figure 4.6).

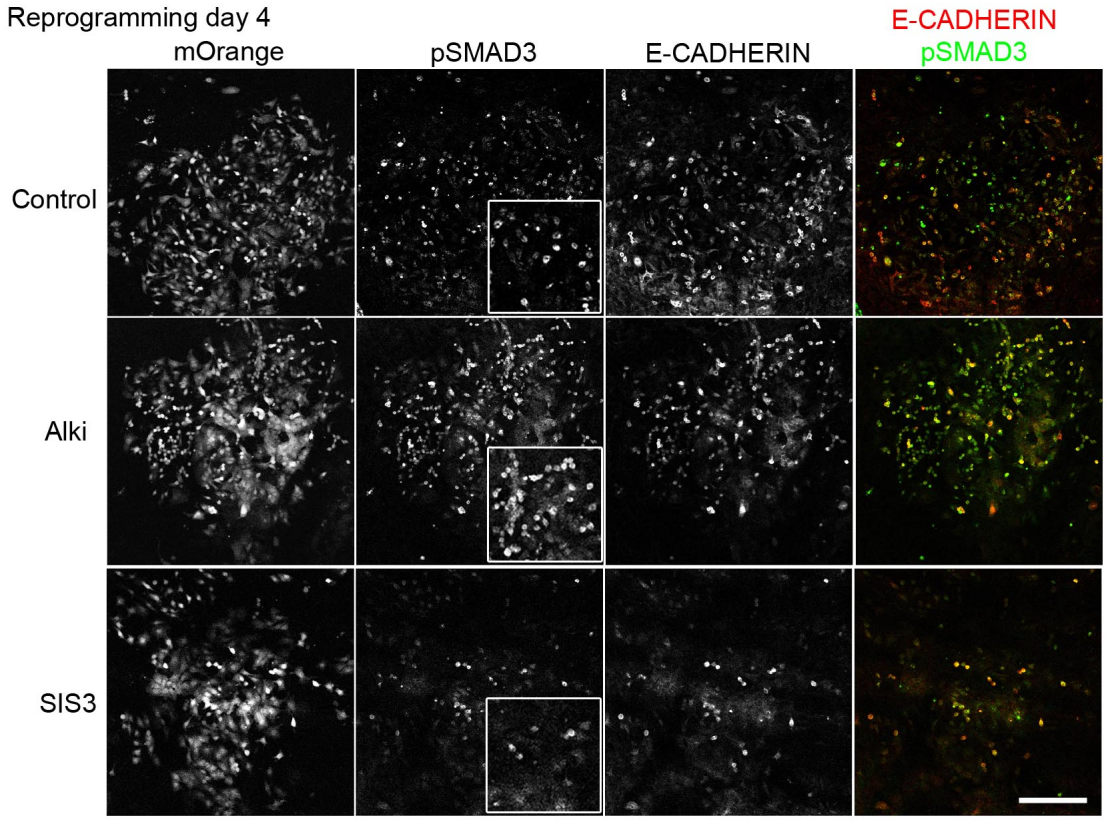

Figure 4.5. Active Smad3 (pSmad3) levels at day 4 of reprogramming

Confocal microscopy at day 4 of reprogramming staining for pSMAD3 and E-CADHERIN. Scale bar represents 100 μM.

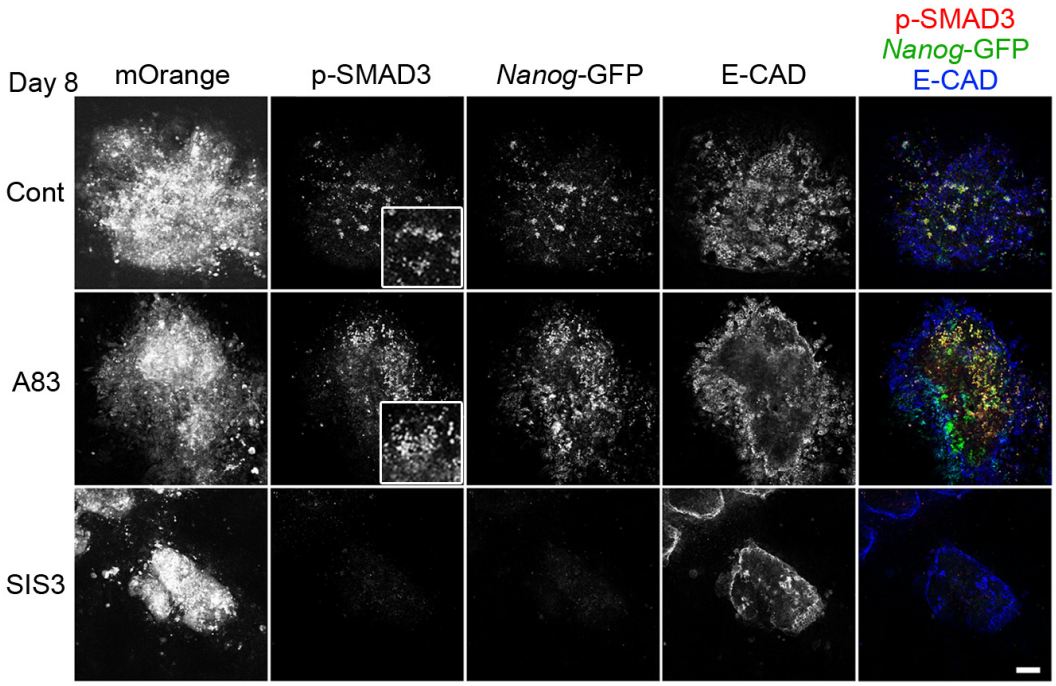

Figure 4.6. pSmad3 levels in colonies at day 8 of reprogramming

Confocal microscopy at day 8 of reprogramming with staining for pSMAD3, E-CADHERIN, and total SMAD3 in combination with a *Nanog*-GFP reporter. Scale bars represent 50 μM.

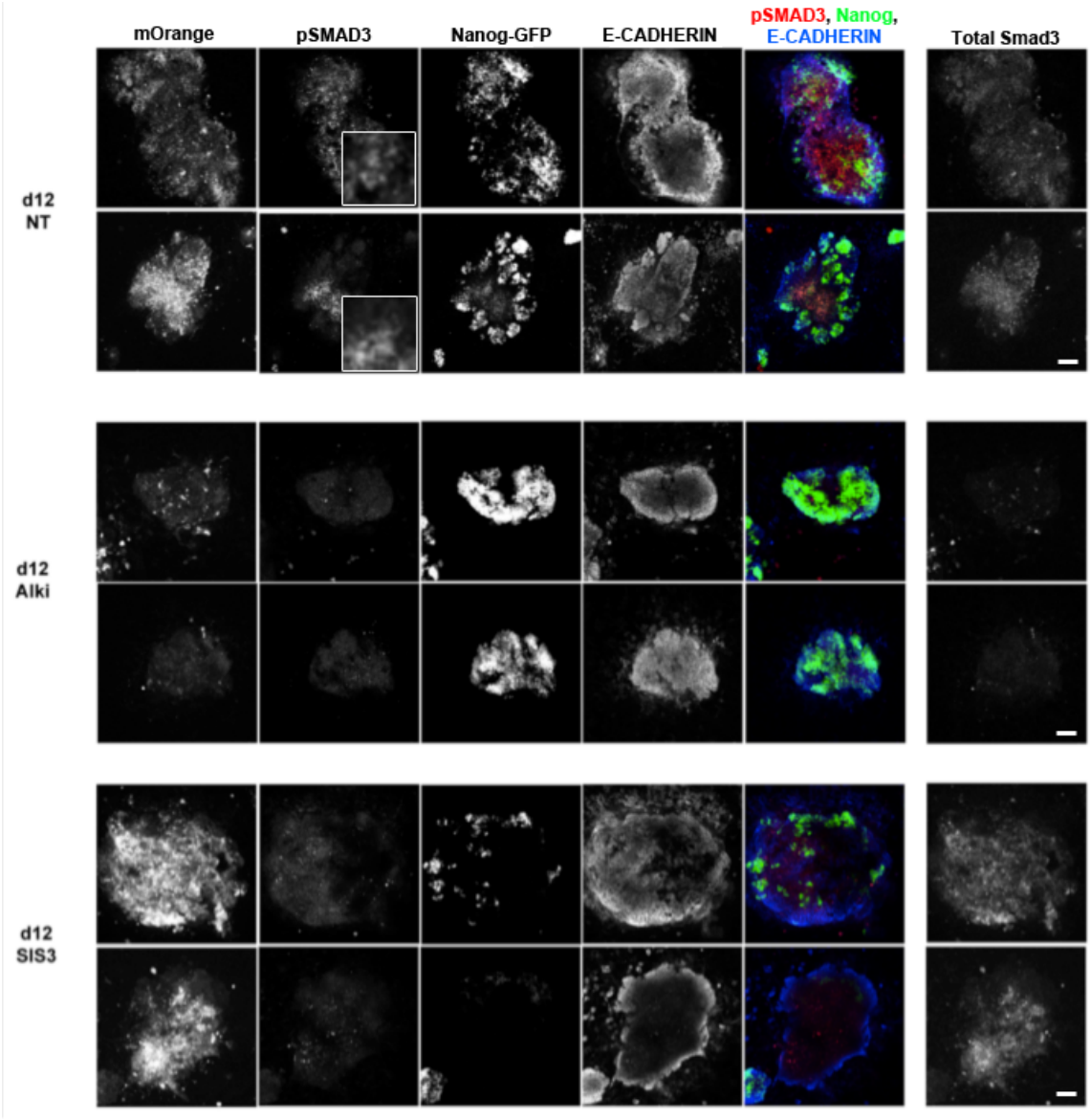

Figure 4.7. Day 12 reprogramming colonies have low levels of pSmad3

Confocal microscopy at day 12 of reprogramming with staining for pSMAD3, E-CADHERIN and SMAD3 in combination with the Nanog-GFP reporter. Scale bars are 100 μM.

To quantify pSmad3 levels during reprogramming in the presence of Alki, a celigo imaging cytometer was used to image entire 6 well plates stained for pSMAD3 at day 4 of reprogramming. Day 4 was chosen because cells are less dense at this stage of reprogramming, allowing for quantification of single cell fluorescence characteristics. Quantifying an entire well, the total mOrange$^+$ cells were first enumerated and then assessed for pSMAD3 levels in cells. A cut off was set such that only cells with punctuated, high nuclear levels of pSMAD3 were counted as positive (Figure 4.8A). Of all mOrange$^+$ cells, ~40% were pSmad3$^+$ at day 4 of reprogramming in the absence of inhibitors, whereas in the presence of the Alki, ~60% were pSmad3$^+$ (Figure 4.8B). Another commonly used TGF-β receptor inhibitor SB431542 (SB43) exhibited similar results, producing higher levels of pSmad3 in cells undergoing reprogramming at day 4 (Figure 4.8B). While these data contradicted what the inhibitors were supposed to do, I could confirm that these inhibitors blocked SMAD2/3 phosphorylation upon 1-hour TGF-β treatment as expected (Figure 4.8C). Similarly, when wt MEFs were cultured in the presence of the inhibitors for 24 hours, I could observe decreased numbers of MEFs with nuclear p-SMAD3 staining, compared to no inhibitor treatment (Fig 4.8D). However, when the culture period was extended to 4 days, I could again observe increased pSMAD2/3 in the presence of either Alki or SB with wt MEFs, similarly to the reprogramming context (4.8E). These results indicated that increased pSMAD2/3 levels in the presence of Alki were not specific to the reprogramming context, but rather dependent on the prolonged culture time in the presence of inhibitors. The TGF-β signalling pathway has auto-regulatory mechanisms, and the chemical inhibitors may not 100% block the receptor kinase activity (Itoh et al., 1998; Shi and Massague, 2003a). A small leak of the receptor kinase activity may be skewing the regulatory mechanisms, and resulting in the unexpected increased pSMAD2/3 levels in the presence of Alki.

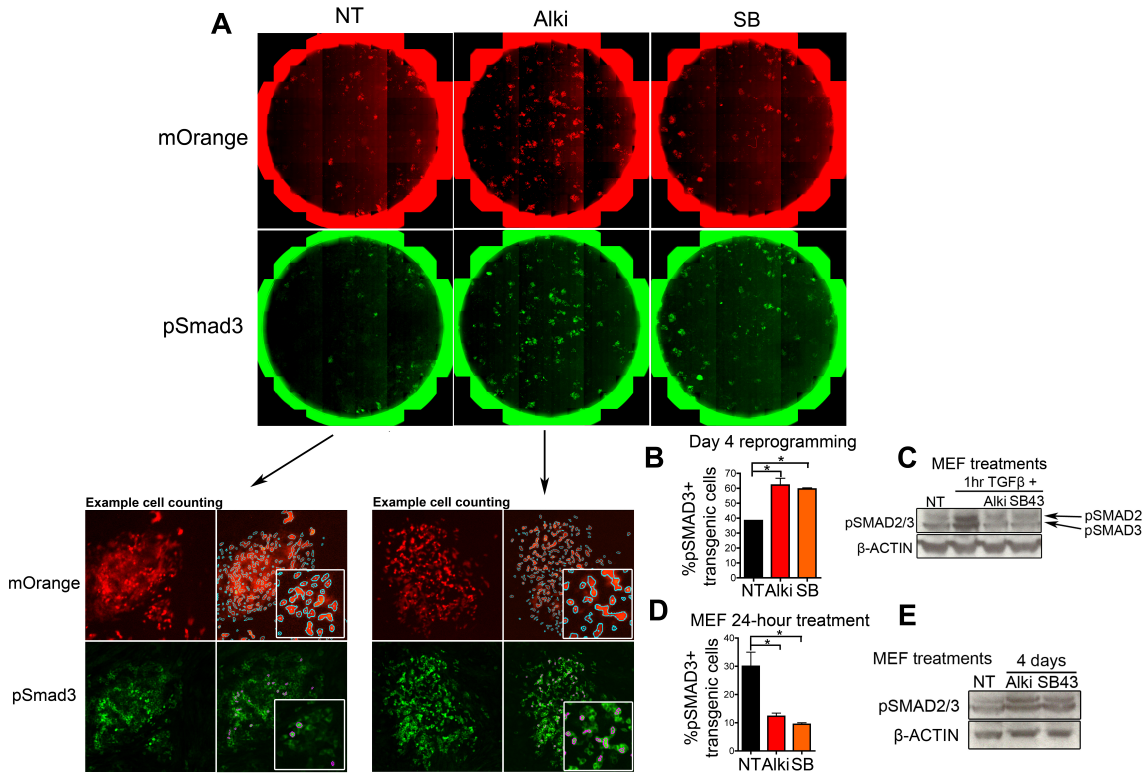

Figure 4.8. Quantifying pSmad3 during reprogramming

A) Celigo imaging cytometer analysis of 6-well plates at day 4 of reprogramming with staining for pSMAD3 and the mOrange reporter for identification of transgene expressing cells. Cells with punctate nuclear staining were considered positive for pSMAD3, as indicated by the lower left panels with cells outlines in blue. B) Graph displays the percent of transgenic cells (mOrange+) positive for pSMAD3 staining, in 3 independent experiments. C) Western Blot displaying pSMAD2/3 levels in MEF treated for 1-hour with TGF-β ligand in combination with Alki (A83-01) or SB43 (SB431542). D) Percent of transgenic cells with nuclear pSMAD3 staining after 24-hour treatment with Alki or SB43. E) Western blot showing pSMAD2/3 after 4-day treatment of MEF with Alki or SB43. Error bars indicate standard deviation (s.d.) from 3 independent experiments with 2 experimental replicates. *P-value <0.05 based on a two-sided t-test.

4.2.2 Constitutively active Smad2/3 boost reprogramming

The results above suggest that active Smad2/3 is concordant with successful reprogramming. To assess if Smad2/3 actively support the reprogramming process, constitutively active (CA) *Smad2CA* and *Smad3CA* were over expressed during reprogramming. The pMXs retroviral packaging system (Kitamura et al., 2003) was used to infect and over express *Smad2CA/3CA* constructs in our TNG-MKOS MEF reprogramming system. The engineered *Smad2CA/3CA* constructs have c-terminal base substitutions producing phosphomimetic pepetides, mimicking their phosphorylated and active state (Chipuk et al., 2002; Funaba and Mathews, 2000). Using the TNG-MKOS MEF reprogramming system, cells were infected with Smad2CA/3CA or control blue fluorescent protein (BFP) retroviral particles and reprogrammed for 12 days. The infection with retroviral particles is mildly toxic to the cells, as can been seen by a reduced reprogramming efficiency in our control condition as compared to previous reprogramming experiments (compare figure 4.2 and 4.9a). There was an over 6-fold increase in *Nanog*-GFP^{+} colonies in the presence of *Smad2CA* or *Smad3CA* (Figure 4.9A). When expressed together, Smad2CA/3CA demonstrated a 10-fold increase in reprogramming efficiency, perhaps as a result of the cooperative nature of Smad2/3 action (Figure 4.9A). Time course FACS analysis revealed that *Nanog*-GFP was activated more robustly in the presence of *Smad2CA/3CA*, while the onset of *E-cadherin* (*E-cad*) expression during MET was not affected (Figure 4.9B and C).

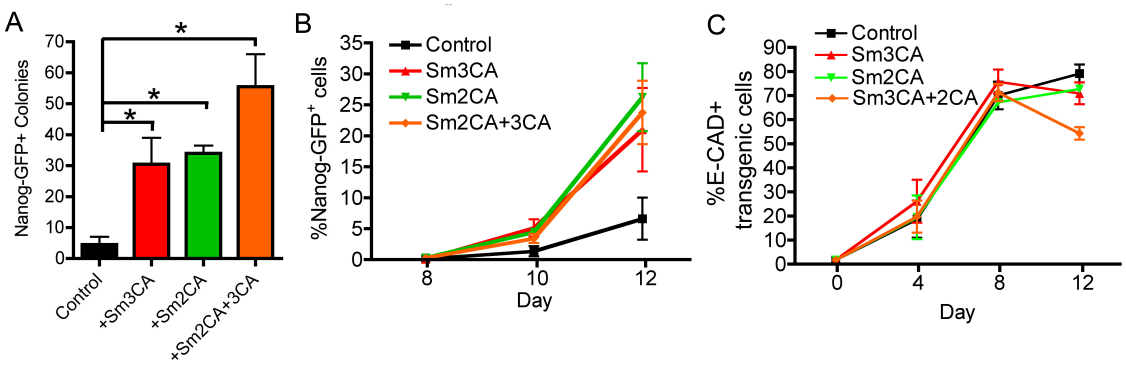

Figure 4.9. Constitutive active Smad2/3 enhance reprogramming

A) Total *Nanog*-GFP+ colonies at day 12 of reprogramming with various treatments. TNG-MKOS MEF were reprogrammed and 'topped-up' with retroviral mediated over-expression of constitutively active *Smad2/3* (*Sm2CA/Sm3CA*) or blue fluorescent protein (BFP; control) B/C) FACS analysis quantifying *Nanog*-GFP/E-CADHERIN (E-CAD) expression in transgenic (mOrange+) cells during reprogramming. Error bars indicate standard deviation (s.d.) from 3 independent experiments with 2 experimental replicates. *P-value <0.05 based on a two-sided t-test.

To investigate the dynamics of reprogramming progression in the presence of *Smad2CA/3CA*, a time-course FACS analysis was performed with the CD44 and ICAM1 surface markers (Figure 4.10). Reprogramming with the 4 Yamanaka factors, plus a combination of Smad2CA, Smad3CA or both expressed together resulted in a similar accelerated reprogramming phenotype. Reprogramming progressed with marked acceleration apparent at day 8 of reprogramming as the intermediate $CD44^{low}$, $ICAM1^{low}$ stage is achieved rapidly when *Smad2CA/3CA* are over-expressed (Figure 4.10 days 4-8). *Nanog*-GFP expression was observed ~5 days earlier with *Smad2CA/3CA* as compared to the control (Day 10 compared to 15, Figure 4.10), in parallel with

accelerated ICAM1 up regulation (Figure 4.10 days 8-12).

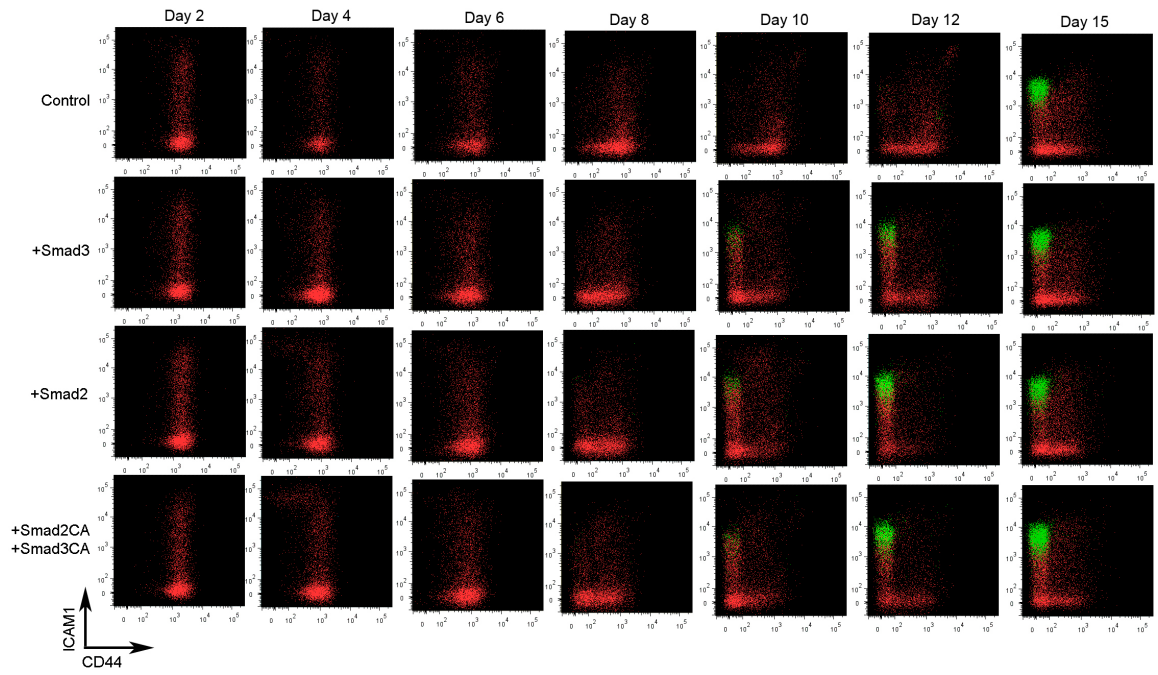

Figure 4.10. Reprogramming progression with Smad2CA/3CA

FACS analysis was performed with staining for ICAM1 and CD44 over a 15-day reprogramming experiment with over-expression of *Smad3CA* or *Smad2CA* or both, compared to control (BFP). FACS plots show only transgenic (mOrange+) cells, green dots indicate *Nanog*-GFP positive cells.

4.2.3 RNA sequencing reveals global acceleration and bypass of aberrant transcriptional programs when reprogramming with addition of Smad3CA

Global gene expression profiling was performed at day 3, 6, 8, and 10, comparing reprogramming with *Smad3CA* and a vector expressing blue fluorescent protein (BFP; control). A list of pluripotency-associated genes was investigated across all time-points of reprogramming, revealing very little expressional changes between *Smad3CA* and control during the first 8 days of reprogramming. However, by day 10, almost every pluripotency gene assessed was expressed higher in the presence of

Smad3CA than control, as confirmed by gene set enrichment analysis (Figure 4.11A and B)

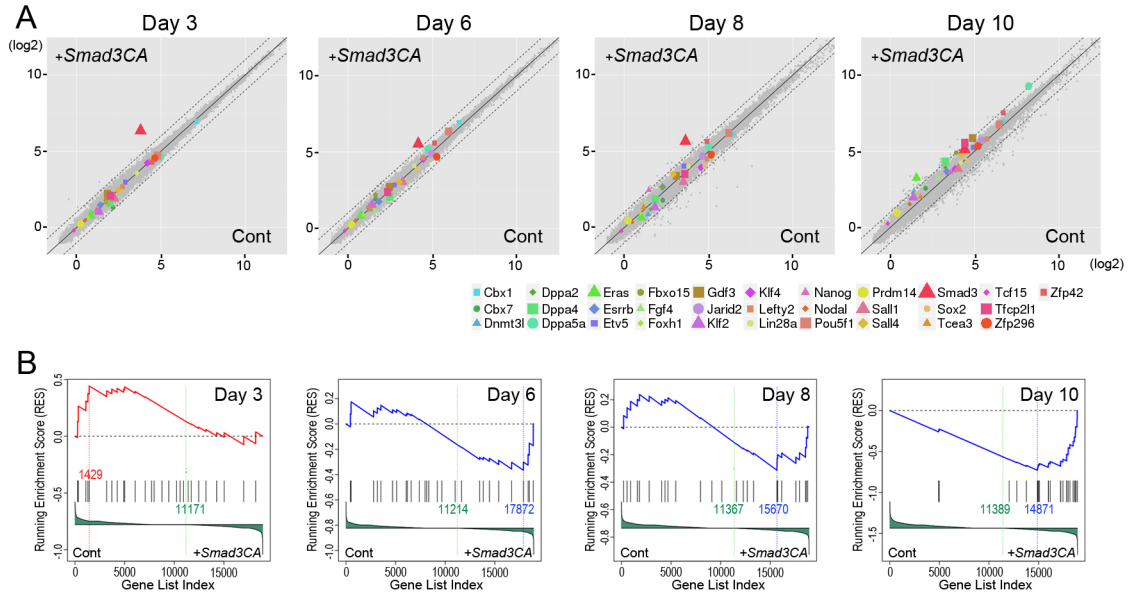

Figure 4.11. Pluripotency gene expression during reprogramming with addition of *Smad3CA*

Global gene expression was assessed with RNA sequencing over a 10-day reprogramming experiment. A) Expression patters of all genes (grey dots) comparing control and Smad3CA samples on day 3, 6, 8, 10 of reprogramming. A panel of 30 pluripotency genes are highlighted with various coloured shapes showing their expression profiles during reprogramming, demonstrating a marked increase in expression of almost all genes in the presence of Smad3CA at day 10. The dashed lines indicate 2-fold differences between the samples. B) Gene Set Enrichment Analysis (GSEA) for the pluripotency genes listed in (A). ES; Enrichment Score, Nom. p-val; nominal p value.

To further explore the contribution of *Smad3CA* during reprogramming we performed gene-clustering analysis, grouping 3750 genes with above background expression levels (mean count >= 2) and low variation (coefficient of variation < 100) into 8 distinct expression patterns over the course of the reprogramming experiment. Seven clusters in control and +*Smad3CA* sample series were very similar, including groups down-regulated by day 3 (Dw), transiently up-regulated at day 3 (D3Up), transiently up-regulated at day 8 (D8Up), highly (D10Up) or slightly (D10sUp) up-regulated at day 10 but low expression in ESCs, expected to be up-regulated after day 10 (Up), and finally a down-then-up cluster (DwUp) (Figures 4.12A). Additionally, the control series had a cluster of transiently up-regulated genes on day 6 (D6Up), and +*Smad3CA* series had a group of genes up-regulated earlier than the Up genes (eUp) (Figures 4.12A). To compare trends across the *Smad3CA* and control treatments, a cord diagram was produced to display where genes within given clusters were located in the other treatment. The cord diagram revealed that genes in Dw, D3Up, Up, and DwUp clusters were very similar between control and *Smad3CA* samples (Figure 4.12B). However, several differences emerged, including with the control D8Up genes, which were classified into *Smad3CA* D10sUp genes (108 genes), indicating this aberrant up-regulation was minimized in the presence of *Smad3CA* (Figure 4.12C). About 24% and 45% of control D10Up genes were cross-classified into D8Up (124 genes) or D10sUp (233 genes) *Smad3CA* clusters, where the peak of the transient up-regulations in *Smad3CA* were accelerated to day 8 or bypassed altogether (Figure 4.12D and E). Most of control D10sUp genes fell into *Smad3CA* D10Up cluster (102 genes) (Figure 4.12F). Finally, almost all genes that belonged to eUp, a unique cluster of early activated genes in the *Smad3CA* series, were classified in the late-activated control Up cluster (120 genes) (Figure 4.12G). These results demonstrate a swifter and more efficient transition to an ES/iPSC like state, when reprogramming with addition of *Smad3CA*.

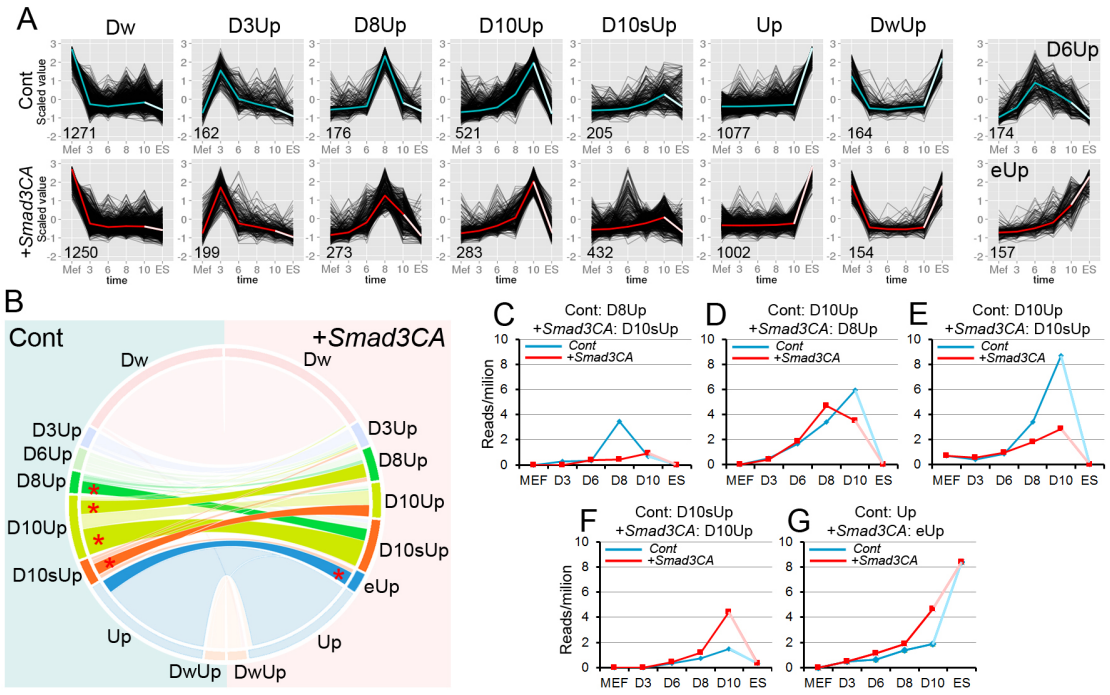

Figure 4.12. Gene expression clustering analysis

A) Gene expression categories clustered by similar gene expression changes during reprogramming with (+*Smad3CA*) and without (Cont) *Smad3CA* over expression. The blue and red lines represent mean values of each cluster. Numbers of genes in each category were indicated. Note values of ESC samples (ES) are included in the analysis to incorporate directions of gene expression changes after day 10. B) A chord diagram to compare distribution of genes classified in each category in control (left) and +*Smad3CA* (right) reprogramming. Asterisks indicate cross-classified gene groups with more than 100 genes. C-G) Expression patterns of the cross-classified gene groups highlighted in (B) in control (blue) and +*Smad3CA* (red) reprogramming. The values represent median RPM.

4.3 Discussion

4.3.1 Smad2/3 are active during reprogramming

Upon TGF-β ligand binding receptors, the intracellular kinase domains of the receptors are activated. While Smad2/3 are primary targets of TGF-β receptor kinase activity, many other cytosolic proteins are phosphorylated and regulated by the receptors (Massague, 2012; Shi and Massague, 2003a). A specific inhibitor of Smad3 (SIS3) was used to assess if blockage of the TGF-β-Smad3 pathway was the major cause of the reprogramming enhancement by the Alki. Surprisingly the SIS3 phenotype suggested that Smad3 activity might be required for successful reprogramming. MEF did not simply die or senesce in the presence of the inhibitor, but appeared to be stuck at an intermediate stage of reprogramming. Subsequent work demonstrated that Smad2/3 are phosphorylated and active during reprogramming, and are active in cells with *Nanog*-GFP expression as reprogramming proceeds. Those results suggest that Smad2/3 might be actively involved in establishing pluripotency network gene activation. Previous work has described roles for Smad2/3 in the specification of various lineages during development, with extensive work demonstrating a role for Smad2/3 during mesenchyme differentiation (Medici et al., 2011; Miettinen et al., 1994; Piek et al., 1999; Valcourt et al., 2005; Xu et al., 2009). For that reason it was surprising to identify that Smad2/3 actively benefit the reversion of fibroblasts to a pluripotent state. However, a hallmark of Smad2/3 signalling is the context dependency, whereby different cellular states direct Smad2/3 transcriptional responses. In the presence of 4 core pluripotency genes being over-expressed, it is not entirely surprising that Smad2/3 would not perform the same role as they do in the maintenance of mesenchymal tissue.

4.3.2 Alki treatment results in increased Smad2/3 activity during reprogramming

Treatment with Alki during reprogramming was expected to decrease Smad2/3 mediated signalling during reprogramming. Conversely, Alki did not reduce, but rather increased, pSMAD2/3 levels. The TGF-β signalling pathway is auto-regulatory whereby

activation of Smad2/3 is a transient response and high levels of activation results in a dampening of TGF-β cell responsiveness, through modulation of receptor activity or ligand expression (Itoh et al., 1998; Shi and Massague, 2003a). For example, within minutes of TGF-β stimulation, SMAD2/3 bind the promoter and activate transcription of the negative regulator Smad7, which directly interacts with TGF-β receptors and inhibits their kinase activity (Kavsak et al., 2000; Suzuki et al., 2002). SMAD7 also induces TGF-β receptor ubiquitination and subsequent degradation (Ebisawa et al., 2001; Kavsak et al., 2000). Similarly another TGF-β negative regulator, *SnoN*, is up regulated within 2-hours of TGF-β stimulation. SNON can directly bind SMAD2/4 and act as a transcriptional repressor (Luo et al., 1999). Thus, inhibiting TGF-β receptors would also alleviate the negative feedback loops of the pathway. It is possible that prolonged exposure to the Alki during reprogramming, and corresponding extended absence of negative feedback loops, primes cells for an eventual increased TGF-β-Smad2/3 response by stabilising TGF-β receptors on the membrane. Indeed results in this chapter demonstrate that a 1-hour treatment of MEF with Alki results in reduced pSmad2/3, and it is only after a prolonged 4-day treatment that an increase in phophorylation levels was observed.

4.3.3 Constitutive active Smad2/3 boost reprogramming

Expression of *Smad2CA* and *Smad3CA* demonstrated that constitutively active versions of those proteins not only boost overall reprogramming efficiency, but also accelerate the activation of the pluripotency network and bypass some aberrant transcriptional programs. Many studies have characterized a role for Smad2/3 in maintenance of pluripotency, and directly regulating the *Nanog* locus in stem cells (Gaarenstroom and Hill, 2014; Sakaki-Yumoto et al., 2013; Xu et al., 2008). Smad2/3 play pivotal roles in maintenance of mouse epiblast like stem cells (EpiSCs) and the transcriptionally similar human ES cells, whereby inhibition of TGF-β signalling results in an exit from self-renewal and pluripotency (Beattie et al., 2005; James et al., 2005; Vallier et al., 2005).

SMAD2/3 interact with many transcriptional activators, silencers and nucleosome remodelers. During development, transcription factors bind to Smad2/3 and recruit them to target loci, whereby SMAD2/3 enable recruitment of transcriptional machinery (Dahle et al., 2010; Estaras et al., 2012; Feng et al., 1998; Germain et al., 2000; Kato et al., 2002; Pouponnot et al., 1998). It is tempting to speculate that Smad2/3 perform a similar role during reprogramming, whereby OCT4, SOX2, KLF4 or C-MYC would recruit them to target loci to activate or silence transcription. Indeed, other groups have demonstrated that OCT4 immunoprecipitates with SMAD3, and directs SMAD3 to its target loci in mouse ES cells (Mullen et al., 2011). In a situation analogous to reprogramming, the over-expression of *MyoD1* in mES cells can recruit SMAD3 to MYOD1 target sites, away from OCT4 bound sites, while it was not sufficient to differentiate ESCs to myoblast in this context. Such data suggests that Smad3 is directed to the target sites of whichever master transcription factor is in greatest abundance, where Smad2/3 recruit the necessary nucleosome remodelers and transcriptional machinery to activate or repress transcription. It is tempting to speculate that during reprogramming, Oct4 and perhaps the other Yamanaka factors, are recruiting Smad2/3 to pluripotency-associated loci to activate transcription.

The RNA-seq results demonstrated that some transiently activated gene clusters in the control conditions were either not activated at all, minimally activated, or their activation was accelerated to earlier time points when reprogramming with addition of *Smad3CA*. Throughout the reprogramming process, the transcript levels for endogenous Smad2/3 are certainly lower than the exogenously expressed Yamanaka factors. It was not assessed how the levels of active Smad2/3 protein compared to the reprogramming factors, but it is possible that over-expressed factors are in far greater abundance than endogenous active Smad2/3. Therefore, most of the expressed OCT4, SOX2, KLF4 and CMYC could be without SMAD2/3-cofactor, which might compromise their ability to activate target genes or result in different transcriptional response at target loci. With *Smad3CA* over expression, and perhaps more potently with Smad2CA/3CA combined, it is likely that the Yamanaka factors can more robustly manipulate transcriptional programs, enabling a more swift and direct reprogramming process.

4.3.4 Bypassing of senescence in the presence of Alki is Smad2/3 independent

The resulting increase in reprogramming efficiency with *Smad2CA/3CA* was not as potent as addition of the Alki, suggesting additional mechanisms of Alki action. The RNA-seq results confirm that transcript levels from the senescence associated CDKN2a ($p19^{ARF}$/$p16^{Ink4a}$) and CDKN1a (p21) loci are not strikingly increased or decreased in the presence of Smad3CA over expression. Thus, the decreased p19 level in the presence of Alki may be a result of other down-stream pathways under TGF-β receptors. TGF-β signalling converges on the senescence program through many unique mechanisms. While Smad2/3 directly regulate $p19^{ARF}$ thereby activating *p53* in fibroblasts, there are many reports of Smad2/3 independent TGF-β senescence responses. For example, knockdown of p38MAPK severely diminishes the TGF-β induced senescence response in MEF (Zheng et al., 2010). Interestingly, p38MAPK directly regulates NF-κB, which activates senescence associated secreted proteins (SASPs), as part of a paracrine senescence program (Freund et al., 2011). SASPs initiate senescence pathways in neighbouring cells, a process that helps establish a robust, localized senescence response in tissues and in cells in culture (Acosta et al., 2013; Freund et al., 2011). It is possible that SASPs are produced during reprogramming, and may represent part of the TGF-β induced p19 senescence response. Intriguingly and unreported in the literature, many of the SASP response genes such as IL6, GROα, MCP-2 and IL-1α are massively up regulated during reprogramming and down-regulated in iPSCs, as can be observed by inspection of time-course RNA-seq data in this thesis and published data sets ((O'Malley et al., 2013; Samavarchi-Tehrani et al., 2010). It is therefore likely that the paracrine senescence response may represent an, as yet unidentified, barrier to reprogramming and that TGF-β ligands may contribute to that response. Future work should explore the potential SASP involvement in reprogramming, as targeting of secreted ligands represents an attractive approach to enhancing reprogramming efficiency.

5: Chapter 5- How Smad3 enhances reprogramming

5.1 Introduction

Extensive work has characterized the role of Smad2/3 transcription factors during maintenance of, and transitions between, cellular states. Often the roles of Smad2/3 are in providing a rapid, tuneable means to boost or repress transcriptional responses. Indeed, short pulses with TGF-β ligands to activate Smad2/3, result in hundreds of genes up or down regulated within 1-2 hours (Bertero et al., 2015; Estaras et al., 2012; Mullen et al., 2011). How Smad2/3 accomplish such vast and rapid transcriptional responses is a result of interaction with a number of transcriptional regulators including CBP/p300 co-activators (Feng et al., 1998), NuRD, N-CoR, Sin3 co-repressor complexes (Akiyoshi et al., 1999; Kim and Lassar, 2003; Luo et al., 1999), SWI/SNF chromatin remodelers (Xi et al., 2008), MLL histone H3 Lys4 (H3K4) methyltransferase complex (Bertero et al., 2015), and Mediator complex (Kato et al., 2002). Work in this thesis demonstrated that over expression of Smad2CA/3CA vastly accelerates the reprogramming process, and results in over 6-fold increase in efficiency. This chapter tests the hypothesis that Smad2CA/3CA boost reprogramming through interaction with chromatin remodelers or transcriptional activators, in cooperation with the exogenous reprogramming factors, to enable establishment of pluripotency.

5.1.1 Aims of this chapter

The following experiments aim to investigate the molecular partners and corresponding functional consequence of Smad3-protein interactions during reprogramming. Co-

Immunoprecipitation experiments were preformed to look for Smad3 interactions with the reprogramming factors as well as transcriptional regulators. To assess if Smad3 is localizing at pluripotency loci, chromatin-immunoprecipitation followed by qPCR (ChIP-qPCR) is performed. Finally, to assess the consequence of Smad3 binding to genomic regions, histone modifications were also assessed by ChIP-seq.

5.2 Results

5.2.1 Molecular interactions of Smad3 during reprogramming

A primary role for Smad2/3 in cells is to drive transcription of target genes. Smad2/3 only weakly bind DNA and require master transcription factors to direct them to target loci, providing a mechanism of contextual regulation (Derynck and Zhang, 2003) (Massague, 2012). To test if the master transcription factors OCT4, SOX2 and KLF4 might interact with SMAD3 during reprogramming, co-immunoprecipitation (Co-IP) followed by western blot analysis was performed. At day 4 of reprogramming, all three master transcription factors co-immnoprecipitated with SMAD3, albeit with a very weak interaction between SMAD3 and KLF4 (Figure 5.1).

SMAD2/3 activate transcription through a number of strategies that include recruitment of nucleosome remodelers like SWI/SNF, histone modifying enzymes such as the MLL complex, and transcriptional machinery proteins like p300 or mediator (Shi and Massague, 2003b). As a broad approach to highlight potential mechanisms of Smad3 action, Co-IP was performed to look for Smad3 bound proteins during reprogramming. The results reveal that Smad3 interacts with the Dpy30(MLL), p300 and Brg1(SWI/SNF), but not Med15(Mediator) at day 4 of reprogramming (Figure 5.1).

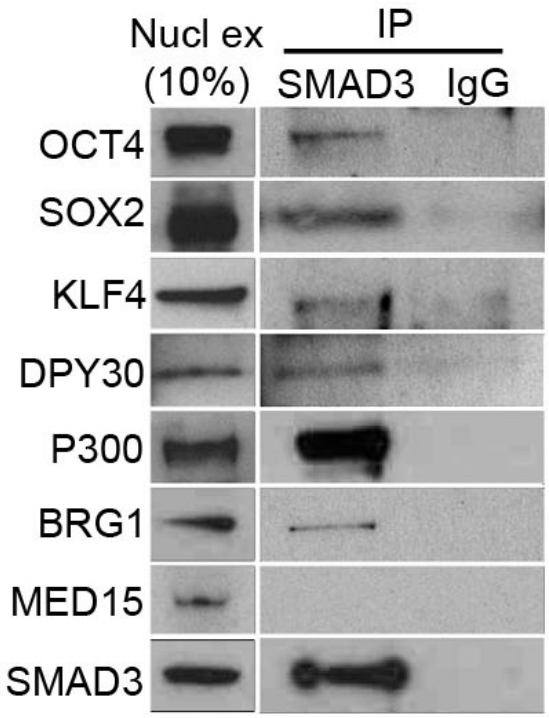

Figure 5.1. Smad3 interacts with reprogramming master transcription factors and nucleosome remodelers

Co-Immunoprecipitation followed by western blot analysis was performed at day 4 of reprogramming. Smad3 antibody or IgG was used to pull-down bound protein complexes. Precipitated complexes were probed with antibodies against the reprogramming master transcription factors as well as a panel of nucleosome remodelers and transcriptional activators. Smad3 pull down was confirmed by staining the blot with a Smad3 antibody raised in a different species.

As a general model, Smad2/3 provide a link between master transcription factors and transcriptional machinery (Massague, 2012; Shi and Massague, 2003b). It is possible that with the over-expression of *Oct4, Sox2* and *Klf4*, that endogenous Smad3 is a limiting factor for transcription machinery recruitment. If Smad3 is a limiting factor,

then over-expression of *Smad3CA* should boost OCT4 interaction with transcriptional machinery. OCT4 immunoprecipitates with DPY30, but not P300, at day 4 of reprogramming in control conditions (Figure 5.2a). When over-expressing *Smad3CA*, there was an apparent increase in OCT4-DPY30 interaction (Figure 5.2a). Quantification of DPY30 band intensity from 2 independent experiments, demonstrated enrichment of OCT4-DPY30 pull-down when over expressing *Smad3CA* (Figure 5.2b).

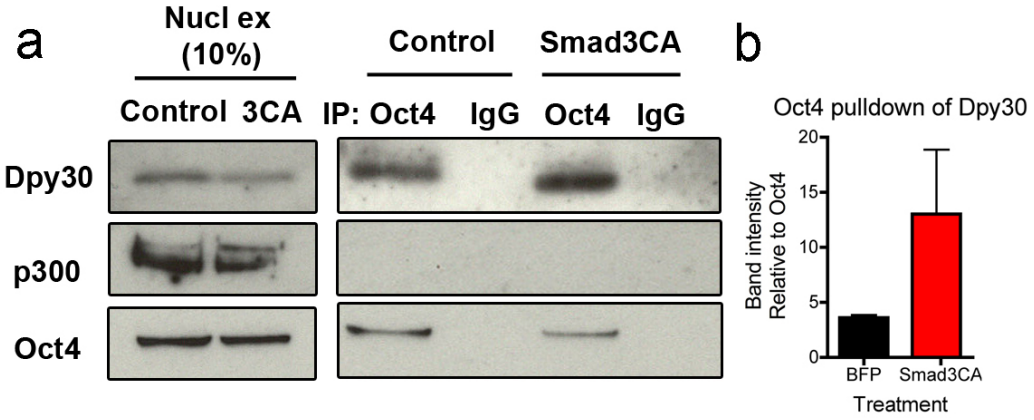

Figure 5.2. Oct4-Dpy30 interaction enhanced with Smad3CA over-expression

A) Co-immunoprecipitation followed by western blot analysis at day 4 of reprogramming. Antibodies against OCT4 or IgG were used to precipitate protein complexes. Complexes were probed with DPY30 or P300 antibodies. Pull-down was confirmed by staining blot with Oct4 antibody raised in a different species. B) Quantification of OCT4 pull-down of DPY30 from 3 independent experiments. ImageJ band intensity was assessed for DPY30 and OCT4 form IP lysate. The intensity of DPY30 band was normalized to Oct4 from the same lysate, to control for differences in pull down efficiency and lysate loading. Error bars indicate standard deviation (s.d.) from 2 independent experiments.

5.2.2 Smad3 chromatin engagement during reprogramming

If Smad3 is important for transcriptional activation of pluripotency associated genes, then it should bind to pluripotency associated loci regulatory regions during reprogramming. Chromatin immunoprecipitation (ChIP) followed by qPCR was performed in ES cells and compared to day 8 of reprogramming. Primers were designed to assess Smad3 occupancy of pluripotency associated genes, based on previously reported binding sites for SMAD3 and OCT4 in ES cells (Chen et al., 2008; Mullen et al., 2011). The results confirmed the exisiting literature that SMAD3 binds many pluripotency-associated loci including *Sall4, Oct4, Fbxo15, Esrrb and Jarid2, Sox2* and *Nanog* in ES cells (Figure 5.3A). SMAD3 binding at day 8 of reprogramming was then examined, revealing that a sub-set of pluripotency loci including *Sall4, Oct4, Fbxo15, Sox, Nanog2 and Gdf3* were enriched (Figure 5.3B). Finally, because Smad3 enhanced interaction between Oct4 and the histone 3 lysine 4 (H3K4) methytransferase complex MLL, H3K4me3 ChIP was performed. H3K4me3 is a histone modification associated with actively transcribed genes (Sandstrom et al., 2014). Smad3 is known to interact with MLL and deposit H3K4me3 at promoters of genes (Bertero et al., 2015). Reprogramming day 8 populations were used for ChIP analysis, revealing H3K4me3 enrichment at pluripotency loci in the Smad3CA treatment as compared to control (Figure5.3C). Interestingly, H3K4me3 marks were enriched even at loci where Smad3 binding was not detected, which may be a result of indirect affects of Smad3 or the limited sensitivity of Smad3 ChIP-qPCR assays. Indeed, a previous ChIPseq study had identified Smad3 binding in ES cells, at all the loci tested in our ChIP experiments (Mullen et al., 2011).

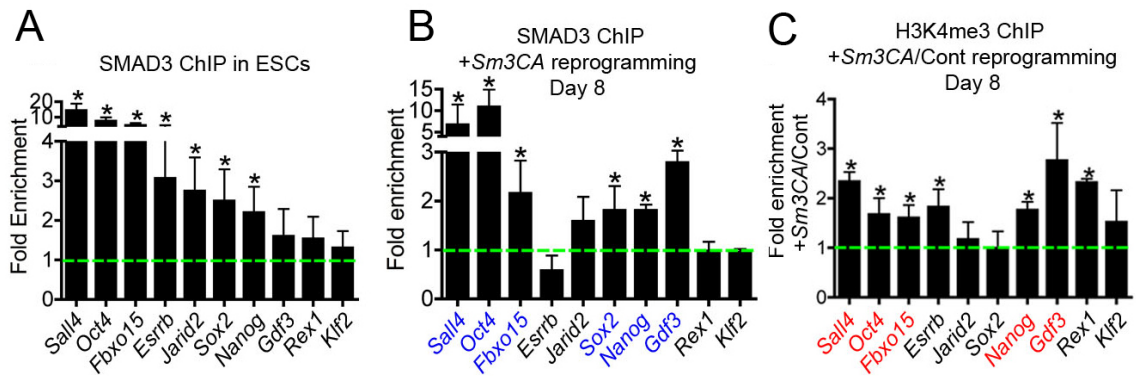

Figure 5.3. Smad3 bind pluripotency loci, and enhances H3K4me3 deposition during reprogramming

A/B) Chromatin immunoprecipitation (ChIP) followed by qPCR was performed in (A) ES cells and at (B) day 8 of reprogramming. Smad3 bound chromatin was precipitated and then qPCR for various pluripotency loci was performed. Day 8 bound genes that were bound in ES cells are shown in blue. Fold enrichment refers to enrichment over genomic regions upstream of *Oct4*, where SMAD3/OCT4 do not bind (Chen et al., 2008; Mullen et al., 2011). C) H3K4me3 ChIP at pluripotency loci transcription start sites, comparing enrichment between Smad3CA and control treatments at day 8 of reprogramming. Genes with H3K4me3 enrichment, that were bound by Smad3 at day 8, are shown in red. Error bars indicate standard deviation (s.d.) from 2(a) or 3(b,c) independent experiments with 2 experimental replicates. *P-value <0.05 based on a two-sided t-test.

5.3 Discussion

5.3.1 Smad3 is a limiting factor for transcriptional activator recruitment

Previous work has extensively demonstrated a role for SMAD3 in the recruitment of transcriptional activators to genomic regulatory regions. One such experiment demonstrated that over expression of constitutively active SMAD3 could boost recruitment of the methylase JMJD3 to neuronal associated loci both *in-vitro* and in the neural tube of the chick embryo (Estaras et al., 2012). Within 30 minutes of TGF-β stimulation in neural stem cells, JMJD3 is recruited to target loci in a Smad3 dependent manner (Estaras et al., 2012). Similarly, in human stem cells, Smad3 and Nanog coordinate the recruitment of Dpy30, as part of the MLL methyltransferase complex, to pluripotency loci to drive transcription (Bertero et al., 2015). A short 2-hour treatment with a TGF-β inhibitor, results in reduction of DPY30 and loss of the H3K4me3 mark at SMAD3-DPY30 target loci, resulting in reduced transcription (Bertero et al., 2015). In the reprogramming context, results presented in this chapter demonstrate that SMAD3 not only interacts with the reprogramming master transcription factors OCT4, SOX2 and to a lesser extent KLF4, but also enhances the interaction between OCT4 and Dpy30, a protein that is part of the histone methyltransferase complex MLL. Those findings suggest SMAD3 not only provides a link between the master transcription factor and transcriptional activators, but also might be a limiting factor in cells, whereby transcription factors compete to recruit SMAD3 and associated proteins to their target loci. Furthermore, it is worth noting that adding more doxycycline, thereby increasing exogenous reprogramming factor expression, does not increase reprogramming efficiency, again suggesting the limiting factors are intrinsic to the cells (unpublished data from our lab).

5.3.2 Smad3 binds pluripotency loci in ES cells and at day 8 of reprogramming

Results presented in this chapter demonstrate that SMAD3 binds pluripotency-associated loci in ES cells and at day 8 of reprogramming. These results suggest SMAD3 is involved in direct regulation of pluripotency loci during reprogramming. Indeed, Chapter 4 results also demonstrated that pluripotency loci in general, have increased levels of transcription by day 10 of reprogramming with Smad3CA. It is tempting to speculate that Smad3 might be recruiting nucleosome remodelers or transcriptional activators, such as p300, MLL complex (Dpy30) or the SWI/SNF complex (Brg1) to bound regions, which results in transcriptional activation. Future work must explore the Smad3 dependent recruitment of such factors to pluripotecny loci.

At day 8 of reprogramming, SMAD3 binding was not detected at all pluripotency loci tested with ChIP-qPCR. The absence of binding could be due to lack of sensitivity of the method, or heterogenous reprogramming populations. Given the low efficiency of reprogramming, even in our advanced reprogramming system, at day 8 the majority of cells will fail to reprogram. Thus, when performing bulk cell analysis, it can be tough to distinguish events important for reprogramming, and those that are not important to the process. However, many of the loci bound by Smad3 at day 8 of reprogramming had enriched H3K4me3 and increased transcriptional activity by day 10 in the RNA-seq data. These data at least suggest a correlation between Smad3 binding and genes which are eventually activated. Alternatively, RNA-seq profiling demonstrated many changes in global transcriptional patterns between Smad3CA and control experiments, revealing that Smad3CA reprogramming bypasses certain aberrant reprogramming transcriptional states. Such results might explain how reprogramming is accelerated by Smad3CA independent of directly binding and regulating pluripotency genes.

6: Supplementary Chapter 6- *Smad3CA* potentiates other master transcription factor mediated cell identity conversions

6.1.1 Note

Work in this thesis lead to the hypothesis that Smad2/3 could be involved in many unique cell identity changes. Collaborating labs performed all of the experiments in this chapter. The results are presented below to demonstrate the impact of the findings from this thesis. All B-cell to macrophage and myoblast to adipocyte conversion experiments were performed by Bruno DiStefano and Tian Tian from Thomas Graf's lab in Barcelona, Spain. Ulrich Pfisterer and Daniella Rylander performed the induced neuron work, from Malin Parmar's lab in Lund University, Sweden.

6.2 Introduction

While induced pluripotent stem cells (iPSCs) have the ability to generate any cell or tissue type, transdifferentiations from one somatic cell type to another are also possible and hold great potential for regenerative medicine. Similarly to iPSCs, transdifferentiation is a cell identity conversion accomplished by over expression of master transcription factors of the desired cell types. The list of possible cell identity conversions is expanding rapidly, ranging from fibroblast to neuron, B-cell to macrophage, myoblast to brown fat, and fibroblast to cardiomyocyte among many others (Graf, 2011). The underlying mechanisms of various transdifferentiation processes have begun to emerge, which ultimately involve the gradual reorganization of chromatin, driven by nucleosome remodelling and histone modifications, establishing a new transcriptional network to generate the final cell type (Di Stefano et al., 2014; Lee et al.,

2014; Tonge et al., 2014; Vierbuchen and Wernig, 2012; Wapinski et al., 2013). Transdifferentiation processes range in duration from a few days, in the case of B-cell to macrophage, to many months for the conversion of fibroblast to mature neuron (Bussmann et al., 2009; Feng et al., 2008; Pereira et al., 2014; Pfisterer et al., 2011). There is almost no information as to why such variances occur; why some cell types are easier to generate than others, and what are the major roadblocks. Furthermore, the question remains if there are shared roadblocks or cellular machinery amongst all forced cell identity change processes.

Recent work illuminated the contextual nature of TGF-β signalling, demonstrating that Smad3 is directed to target loci by cell type specific master transcription factors (Mullen et al., 2011). Chromatin immunoprecipitation (ChIP) experiments revealed that SMAD3 binds entirely unique regions across the genome in a given cell type, co-localizing with OCT4, SOX2 and NANOG in ES cells, PU.1 in pro-B cells, and MYOD1 in Myotubes. Importantly, expression of MyoD1 in ES cells directed SMAD3 to bind MYOD1 binding sites, whilst SMAD3 remained bound to OCT4 bound regions. Those results reveal how transcription factors might compete to commandeer Smad3 to their target loci to activate transcription and force cell identity changes. The master transcription factors that drive these cell identity conversions rely upon cofactors to recruit nucleosome remodelers and transcriptional machinery to target loci. Here we hypothesized that different forced cell identity conversion systems, with unique sets of master transcription factors, might rely upon Smad2/3 to facilitate the conversion process.

Here we studied the forced cell identity conversion of B cells to macrophages, myoblasts to adipocytes, and fibroblasts the neurons with unique sets of master transcription factors and addition of *Smad2CA/3CA*. B cells can be converted to the developmentally related macrophage by expression of a single transcription factor, C/EBPα or C/EBPβ (Xie et al., 2004). Within just 5 days of C/EBPα expression in B cell progenitors, up to 80% of cells have completely down-regulated the B-cell transcriptional program including the surface antigen CD19, and up-regulated the macrophage specific transcription network including the Mac-1 surface receptor.

Myoblasts can also be converted to adipocytes (brown fat) upon over expression of C/EBPβ and PRDM16. Brown fat cells and myoblasts arise from a common progenitor, whereby fate determination is regulated by the transcription factor PRDM16 (Kajimura et al., 2009). Loss of PRDM16 *in-vitro* in precursor cells result in loss of brown fat populations, favouring myogenic lineages, and ectopic expression of PRDM16 in C2C12 myoblasts drives conversion to a brown fat cell fate (Seale et al., 2009). It was subsequently demonstrated that PRDM16 requires endogenous C/EBPβ to initiate the conversion, and over expression of C/EBPβ could not only enhance the process but enable conversion of fibroblasts to brown fat (Kajimura et al., 2009). Finally, in 2010, it was discovered that human fibroblasts could be converted to neurons by over-expression of just 3 neuronal transcription factors: Ascl1, Brn2 and Myt1l (Vierbuchen et al., 2010). The resultant induced neurons (iNs) exhibit typical neuronal morphology with axonal outgrowths, express neuronal genes, fire action potentials and generate functional synapses (Vierbuchen et al., 2010). It was subsequently identified that addition of NeuroD1 could boost the conversion process and maturity of resultant neurons (Pang et al., 2011). Because Smad2/3 interact with many cell-type specific master transcription factors, we tested if Smad2/3 could boost diverse cell identity conversion systems in combination with other cell type specific master transcription factors.

6.2.1 Aims of this chapter

This chapter tests the hypothesis that Smad2/3 are shared underlying cellular machinery involved in diverse cell identity conversions with master transcription factors. The contribution of Smad3 to the processes of converting B-cells to macrophages, myoblasts to adipocytes, and fibroblasts to neurons are investigated. The conversion processes are accomplished by over expression of cell type specific master transcription factors, in various starting cells including from mouse and human origin, with addition of *Smad2CA* or *Smad3CA*. Functional readout varies depending on converted cell types, but involves microscopy, FACS analysis, and electrophysiology with patch-clamp

analysis. The results present an assessment of conversion efficiency and maturity of the converted cell types.

6.3 Results

6.3.1 Smad2CA/3CA boost the conversion of B-cell to Macrophage with C/EBPα

To test if Smad2/3 could boost the conversion of B cells to macrophages, B cell progenitors were harvested from C57Bl/6J mice using magnetic beads selection to capture CD19 positive cells, as previously described (Xie et al., 2004). Retroviral vectors were used to express *C/ebpα* in combination with *Smad3CA* or *Smad2CA*, and transdifferentiation progression was monitored with CD19 and MAC-1 expression by FACS analysis. Within 6 days of *C/ebpα* expression alone, 99% of starting pre-B cells down regulated CD19 expression and gained the macrophage marker Mac-1 expression (Figure 6.1). With the addition of *Smad3CA* or *Smad2CA* expression, the up regulation of Mac-1 expression was significantly accelerated, with ~2-fold increase in MAC-1 expression at day 3. Interestingly, the down regulation of CD19 was unaffected by *Smad2CA* or *Smad3CA* expression (figure 6.1B and C).

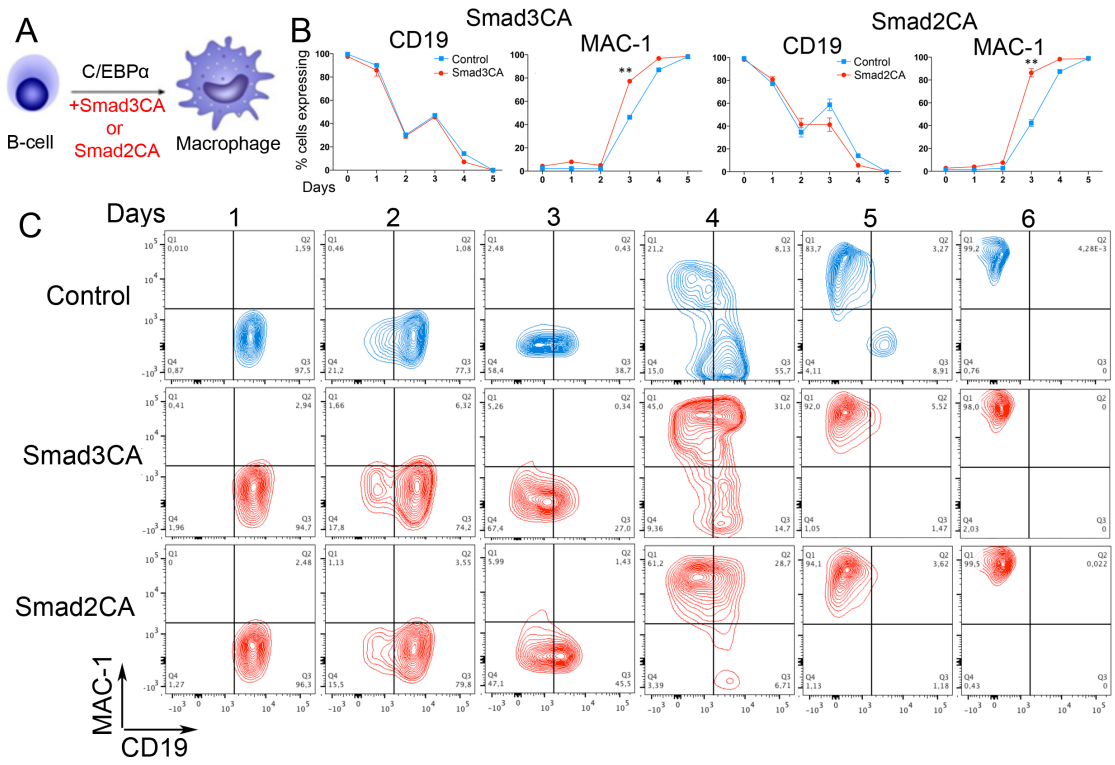

Figure 6.1. Conversion of B cell to macrophage with C/EBPα is enhanced with Smad2CA/3CA.

A) B cells were converted to macrophages by retroviral-transduction of C/EBPα with addition of Smad2CA or Smad3CA. B/C) FACS analysis monitoring the B cell surface marker CD19 and macrophage marker MAC-1 expression during the 6-day transdifferentiation process. Graph values represent average of 2 independent experiments. **P-value <0.01 based on a two-sided t-test.

6.3.2 Myoblast to adipocyte transdifferentiation with PRDM16 and C/EBPβ is enhanced by SMAD2CA/3CA

To examine if Smad2/3 can boost myoblast to adipocyte cell conversion, *Prdm16* and *C/ebpβ* were transduced together with *Smad2CA* or *Smad3CA* into C2C12 myoblasts (Figure 6.2A). After 6 days of induction, wells were stained with Oil Red O,

which stains the lipids and triglycerides within fat cells. In the presence of *Smad2CA* or *Smad3CA*, adipogenesis conversion was boosted, showing a ~20% increase in entire well imaging of Oil Red O (OD 500nm) by day 6 (Figure 6.2B, C and D).

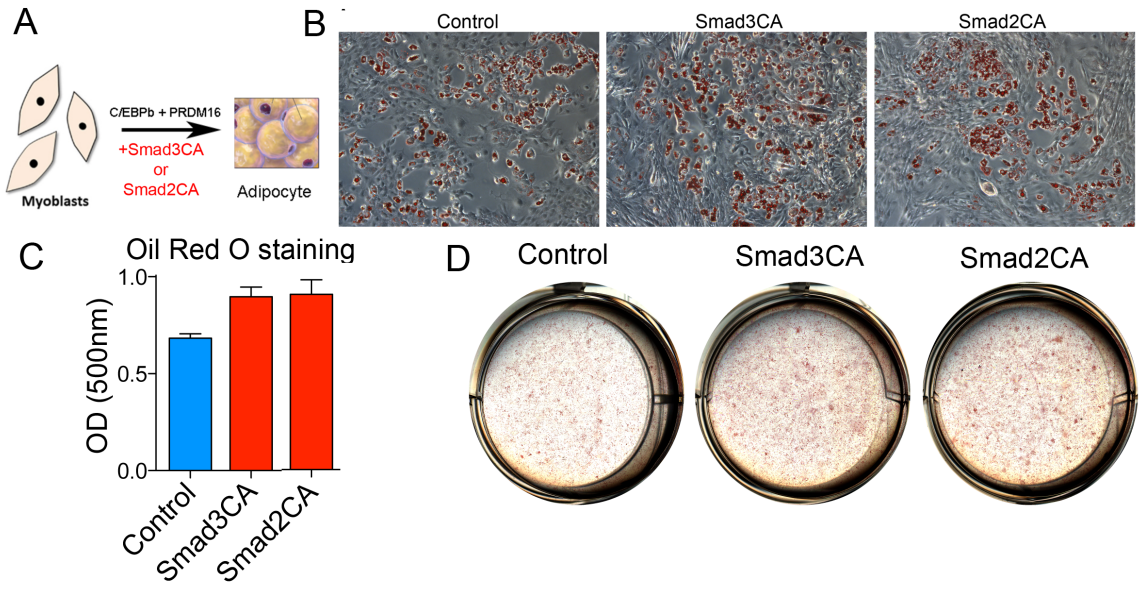

Figure 6.2. Smad2CA/3CA boost myoblast to adipocyte transdifferentiation.

A) C2C12 myoblasts are converted to adipocytes with expression of *C/ebpβ* and *Prdm16*, with addition of *Smad2CA* or *Smad3CA*. B) Oil Red O staining of cells 6 days post induction of transgenes. C) Quantification of Oil Red O images. D) Examples of entire wells stained with Oil Red O. Graphs represent 3 averages of 2 independent experiments with 3 replicates. **P-value <0.01 based on a two-sided t-test.

6.3.3 Induced neuron maturity is increased when generated with addition of *Smad2CA/3CA*

The HFL1 Human fibroblast line was transduced with lentiviral expression vectors encoding *Ascl1, Brn2, Myt1l* and *NeuroD1* (ABMN) with addition of *Smad2CA*

or *Smad3CA* (Figure 6.3A). Neuronal purity of transdifferentiated cultures was assessed using NCAM staining at day 25, revealing no difference between control (ABMN) and addition of Smad2CA or Smad3CA (Figure 6.3B). To assess neuron functional maturity, whole cell patch clamp analysis was performed at day 25, a time at which NCAM positive cells hare maximal and no new neurons are generated. Electrophysiology recordings revealed that control neurons, generated with only ABMN, did not have the ability to fire action potentials (Figure 6.3C and D.) at this early time point, consistent with previous reports {Pfisterer, 2011 #837}. With addition of Smad3CA or Smad2CA to ABMN conversions, a remarkable increase in neuronal maturity was observed, with 66% and 78% of neurons able to fire action potentials, of which 27% and 33% were able to fire repetitive action potentials respectively (Figure 6.3 C and D). Moreover, strikingly, 36% of neurons generated with addition of Smad3CA demonstrated spontaneous action potential activity at day 25, a feature of highly mature neurons able to send and receive synaptic inputs, which are typically only observed after 90 days of conversion (Figure 6.3D) (Pereira et al., 2014).

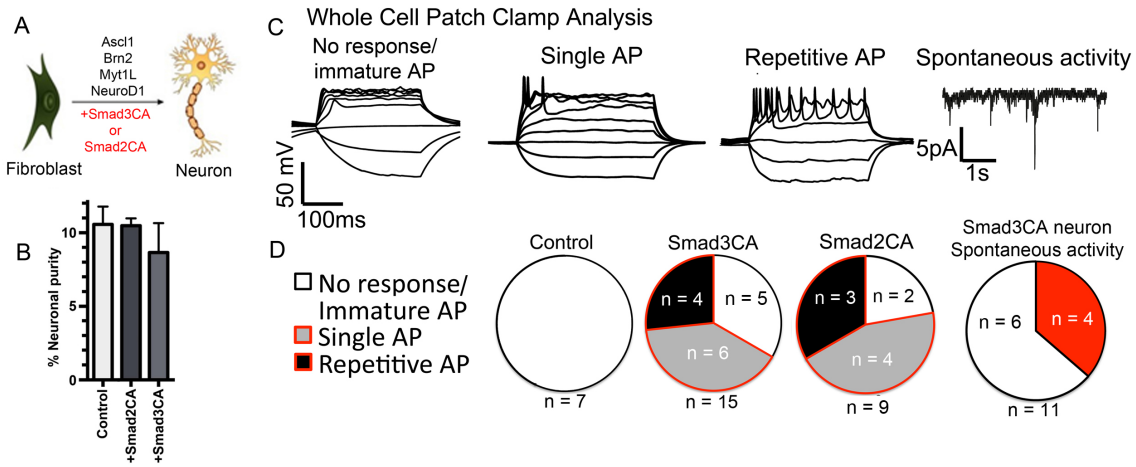

Figure 6.3. Smad2CA/3CA enhances induced neuron maturity

A) Human HFL1 fibroblasts are converted to induced neurons by lentiviral transduction of *Ascl1, Brn2, Myt1L* and *NeuroD1*, with addition of *Smad2CA/3CA*. B) FACS analysis quantifying percent of NCAM+ cells (neuronal purity) at day 25 of the conversion process. C) Representative whole cell patch clamp traces of membrane sodium and potassium currents following induced action potentials. Far right demonstrates spontaneous synaptic activity D) Electrophysiology characteristics of cells recorded at day 25 of induction. Error bars represent standard deviation form 2 independent experiments with 2 technical replicates.

6.4 Discussion

6.4.1 Smad2/3 potentiate diverse transdifferentiation processes

It was possible that Smad2/3's role in induced pluripotency was limited to reprogramming and cell identity conversions from fibroblasts. This chapter explored the ability of Smad2/3 to boost transdifferentiation processes, including from cells developmentally distinct from fibroblasts. The results demonstrate that Smad2/3 can act in many diverse cell types, and boost the conversion of distinct transdifferentiation processes in combination with various master transcription factors (Figure 6.4). It is possible that Smad3 (or Smad2) may interact with the master transcription factors of macrophages, brown fat cells or neurons, in a similar fashion as during induced pluripotency, where Smad3 interacts with Oct4 and boosts its interaction with the MLL histone modifying complex (Figure 5.2 and 6.1). Interestingly, interaction between the pre-B cell transcription factor PU.1 and Smad3 has been previously demonstrated (Mullen et al., 2011). Smad3 and PU.1 bind overlapping genomic loci in pre-B cells and PU.1 knockdown results in reduction of Smad3 occupancy at sites bound by PU.1. Those results suggested that Smad3 was recruited by PU.1 to target loci, analogously to Smad3 recruitment to pluripotency sites by Oct4. Given PU.1 is necessary for transdifferentiation of pre-B cells to macrophages, and Smad3 can boost the process, it would be interesting to investigate PU.1-Smad3 interaction during the conversion.

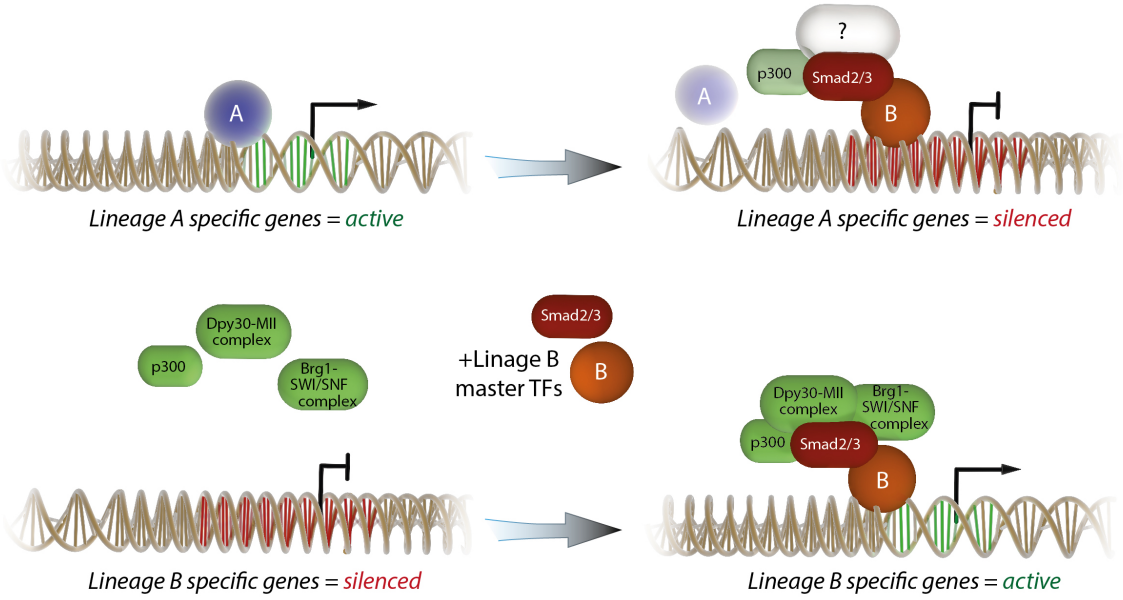

Figure 6.4. SMAD2/3 potentiate cell identity conversions with master transcription factors

Cell identity conversions induced by over expression of master transcription factors (TFs) involve epigenetic and transcriptional changes throughout the conversion process. Smad2/3 interact with many different cell-type specific master TFs, as well as coactivators/corepressors, such as p300, MLL complex and the SWI/SNF complex. During cell identity conversion, Smad2/3 provide a bridge, linking master TFs with various cofactors, which potentiate the epigenetic and transcriptional changes.

6.4.2 Smad2/3 promote neuronal maturity

While the discovery of B cell to macrophage and myoblast to brown fat transdifferentiation were remarkable feats of cellular engineering, it still remained an open question as to the limitations of cell fate conversions. For example, the question remained if cells could be transdifferentiated to cells of a different developmental germ

layer. Upon the discovery of iPSCs, demonstrating a remarkable reversion of fibroblast cell fate, it seemed any cell fate conversions was possible given the right set of transcription factors. Perhaps the earliest and most concrete demonstration of transdifferentiation between different developmental germ layers was achieved when fibroblasts were converted to neurons (Vierbuchen et al., 2010). The finding that Smad3CA can boost the conversion by enhancing neuronal maturity, accelerating the process by months, was a very exciting finding. It was a powerful discovery, demonstrating Smad3 can boost conversions involving different developmental germ layers, further supporting a more ubiquitous role for Smad2/3 is many cell fate changes.

Whole cell patch clamp analysis is a technique commonly used to assess a hallmark of neuronal activity: the ability to control ionic passage and transmit electrical signals across the membrane (Karmazinova and Lacinova, 2010). One can measure both single and repetitive action potentials from individual neurons using the whole cell patch clamp approach. Whilst Smad2/3 did not enhance the efficiency of converting fibroblasts to neurons, the induced neurons demonstrated striking maturity when converted with the addition of *Smad2CA* or *Smad3CA*. Whole cell patch clamp analysis revealed that, only with the addition of *Smad2CA* or *Smad3CA*, induced neurons fired action potentials at day 25 of conversion. Furthermore, roughly one third of neurons produced with addition of *Smad2CA/3CA* fired repetitive action potentials, a trait indicative of advanced maturity (Pereira et al., 2014). Recent work has highlighted important roles for Smad2/3 during neuronal differentiation, suggesting that Smad2/3 are major determinants of the neural developmental program (Estaras et al., 2012; Garcia-Campmany and Marti, 2007; He et al., 2014; Jia et al., 2009). Perhaps the best example of Smad3 activity was demonstrated by over expression of *Smad3CA* in the chick spinal cord, which resulted in rapid down regulation of neural precursor genes such as *Id*'s and *Sox2*, and activation of neural commitment genes such as *NeuroM, HuCD* and *Tuj1* (Estaras et al., 2012; Garcia-Campmany and Marti, 2007). SMAD3 recruits the H3K27me3 histone demethylase JMJD3 to neural promoters, which produces a transcriptional permissive state (Estaras et al., 2012). Similar roles for Jmjd3 have also been demonstrated in neural stem cells of the subventricular zone, whereby

Jmjd3 is required for neurogensis in adult mice (Park et al., 2014). Previous findings are therefore in agreement with work from this thesis, demonstrating a crucial role for Smad3 in directing neuronal maturation. It is tempting to speculate that Smad3, in conjunction with histone remodelers such as Jmjd3, may orchestrate chromatin remodelling during conversion of fibroblasts to neurons.

7: General Discussion

7.1 General contributions of this thesis

The ability to reprogram somatic cells to a pluripotent state holds tremendous promise for regenerative medicine. One of the more directly applicable findings from this work is the increased efficiency and acceleration of the generation of iPSCs with Smad2CA/3CA. Improving the efficiency and accelerating reprogramming will allow ease of generation of iPSCs and will make the technology more attractive for potential clinical uses, as well as large scale disease modelling approaches, where time and financial burden are of major consideration. Improving the efficiency of fibroblast reprogramming may also benefit the reprogramming of other cells types that may have more direct clinical relevance, such as the reprogramming of peripheral blood, which is an attractive alternative to fibroblasts due to the ease of cell extraction (Staerk et al., 2010). Excitingly, work from this thesis has lead to a collaborative project with the company Roslin Cells, based in Edinburgh, to test Smad2CA/3CA's ability to boost reprogramming of peripheral blood in hopes of making the technology more robust for clinical applications.

One caveat of this work is that the quality or stability of iPSC generated in the presence of Smad2CA/3CA was not assessed. The use of the Nanog-GFP reporter is a valuable tool to suggest the acquisition of a pluripotent like state, but does not suffice to prove the resultant cells are bona fide iPSCs. While the global transcription profiles of resultant iPSCs in the presence of Smad3CA are very similar to ESCs, a more comprehensive battery of pluripotency tests is required. For example, the ability of the cells to give rise to teratomas upon subcutaneous injection into mice, contribute to

chimeras, integrate into blastocysts, and be transmitted through the germline should be assessed. We are currently performing such characterizations.

Years of work exploring TGF-β signalling in embryogenesis have highlighted roles for Smad2/3 in differentiation and specification of the mesendoderm germ lineage. Such seminal works were the foundation for the hypothesis that TGF-β inhibitor with Alki worked to alleviate mesenchymal signalling during the reprogramming of fibroblasts (Ichida et al., 2009a; Li et al., 2010a; Polo and Hochedlinger, 2010). However, while it is clear that TGF-β and Smad2/3 are necessary for primitive streak formation and driving the mesenchymal transcriptional program, conditional knockout/rescues and chimeric embryos have highlighted that TGF-β-Smad2/3 knockouts exhibit phenotypes across all three embryonic germ layers throughout development. Striking examples of conditional or partial *Smad2/3* knockout phenotypes include the malformation of embryonic anterior neural patterning, endoderm defects giving rise to liver failure, and onset of colorectal cancers with near 100% penetrance (Vincent et al., 2003; Weinstein et al., 2001; Yang et al., 1999; Zhu et al., 1998). To date, there is tenuous evidence of TGF-β inhibition contributing to down-regulation of mesenchymal signalling during reprogramming. Experiments in this thesis further demonstrate that transition to the epithelial program with *E-cadherin* expression, is not enhanced by TGF-β inhibition. Finally, over expression of *Smad2CA/3CA* during reprogramming did not increase the mesenchymal gene program, but rather boosted acquisition of pluripotency. It is clear that TGF-β-Smad2/3 signalling is not mesenchyme specific and this thesis further clarifies a model whereby TGF-β-Smad2/3 signalling is a product of the current state of the cell, acting more to potentiate, rather than direct, a given transcriptional program.

Recent work describes essential roles for the TGF-β receptor Alk5, as well as Smad3, in accelerating neuronal differentiation from neural stem cells in both neural tube development and adult stem cell populations (Estaras et al., 2012; Garcia-Campmany and Marti, 2007; He et al., 2014; Jia et al., 2009). Work performed in this thesis describes the accelerated maturation of induced neurons with the addition of *Smad2CA/3CA*. Production of mature neurons, with the ability to fire action potentials

and integrate into neuronal networks, has long been a challenge in the neuron differentiation field (Chambers et al., 2009; Kriks et al., 2011; Miller et al., 2013; Osakada et al., 2009; Saha and Jaenisch, 2009). The strikingly advanced maturity of induced neurons produced with addition of *Smad2CA/3CA* has initiated a new project, investigating neuronal maturation *in vitro and in vivo*, in collaboration with Malin Parmar's group in Lund. The birth of such projects illuminates how the study of basic underlying mechanisms of induced pluripotency can transition to translational applications.

This work has demonstrated that Smad2/3 are not only shared underlying cellular machinery for reprogramming, but also a powerful tool that can potentiate master transcription factor mediated cell identity conversions. Such findings provide the groundwork for exploring cellular identity changes on the basis of what the shared roadblocks are and how we can manipulate cells to be more permissive to any cell identity conversion. The overarching process of rearranging chromatin and transcriptional networks to change cell identity, is a shared feature of all cell identity changes to some degree (Ho and Crabtree, 2010; Sisakhtnezhad and Matin, 2012). Furthermore, certain chromatin states with specific histone modifications, are associated with transcription factor accessibility and tightly correlate with transcriptional activation or repression (Papp and Plath, 2013; Soufi et al., 2015; Wapinski et al., 2013). It was therefore not entirely surprising to find chromatin remodelling cofactors like Smad2/3 to be involved in multiple cell identity conversions. The finding that Smad2/3 might be limiting factors for transcriptional activation further suggests a model whereby transcription factor activity is limited by availability of chromatin remodelling factors and transcriptional machinery. It will be interesting to identify what other major 'bottleneck factors' exist, and if they can be manipulated like Smad2/3 to boost the production of more challenging cell types. It is likely that future work exploring commonalities amongst forced cell identity changes will begin with examination of chromatin remodelling factors.

7.2 Future Directions

7.2.1 Characterizing TGF-β initiated senescence during reprogramming

Work in this thesis revealed a bypassing of p19-induced senescence during reprogramming in the presence of the TGF-β inhibitor Alki. As discussed in chapter 4, TGF-β directly activates p38MAPK, which is linked to activation of $p19^{ARF}$ through mechanisms that likely involve regulation of senescence associated secreted proteins (SASPs) (Zheng et al., 2010). P38MAPK is involved in regulation of NF-κB, which regulates SASPs as part of the paracrine senescence program (Freund et al., 2011). Indeed, reports have demonstrated TGF-β ligands induce SASP responses in neighbouring cells (Hoare and Narita, 2013; Hubackova et al., 2012). Probing 3 independent reprogramming data sets revealed that SASPs are robustly up regulated during reprogramming, representing a possible undiscovered reprogramming roadblock. Future work should involve investigating the SASP response during reprogramming and possible link to TGF-β signalling.

7.2.2 Deepening our understanding of Smad2/3 activation of pluripotency associated genes

Future work must clearly demonstrate that SMAD2/3 are recruiting nucleosome remodelers, histone modifiers or other transcriptional machinery to target loci in the reprogramming context. It is possible the accelerated activation of pluripotency genes in the presence of Smad2CA/3CA during reprogramming could be a secondary effect. While this thesis revealed Smad3 binds pluripotency-associated loci during reprogramming, to clearly demonstrate an effect, it is necessary to show Smad2/3 dependent recruitment of nucleosome remodelers or transcriptional regulators to those sites. One such experiment could involve a short burst of *Smad3CA* expression, followed by ChIP analysis. Preliminary work has demonstrated that activation of dox inducible *Smad3CA* at day 6 of reprogramming, gave rise to substantially more *Nanog*-GFP^{+} colonies just 48-hours later. Thus, it is clear that Smad3CA can have effects on

pluripotency loci such as *Nanog* over short duration of expression, providing a means to test the direct effects of the protein during reprogramming. ChIP analysis can then assess if nucleosome remodelers or transcriptional activator candidates bind pluripotency loci more robustly with short pulses of Smad3CA expression. Candidates should include all of which co-immunoprecipitated with SMAD3 at day 4 of reprogramming, such as P300, MLL (DPY30) and SWI/SNF (BRG1) complexes. Following those experiments, it might be interesting to examine histone modifications that result from such recruitment, and there association with transcriptional activation.

As a complementary approach, an assay for transposon accessible chromatin using sequencing (ATAC-seq) has been undertaken to assess chromatin dynamics during reprogramming with or without *Smad3CA* over-expression. The recently described technique utilizes the TN5 tranposase to insert sequencing adaptors throughout the genome, whereby nucleosome free DNA regions tend to have more adaptor insertions (Buenrostro et al., 2013). Upon sequencing, one can visualize the occupancy of nucleosomes across the genome by deductive analysis, whereby nucleosome rich regions will have fewer sequencing reads. The approach enables the assessment of global chromatin status at an extraordinary resolution (~10bp). I have performed ATAC-seq over a 15 day reprogramming experiment, every 2 days, which will enable the assessment of global chromatin changes elicited by Smad3CA. These results should reveal the nature of Smad3 chromatin reorganization, and further our understanding of Smad3 contribution to cell identity changes. Furthermore, this will be the first experiment to probe global chromatin architecture in such fine detail throughout a reprogramming experiment, which will be of tremendous benefit to understand how gross chromatin architecture in various genomic regions influence the outcome of reprogramming, and how it might be reorganized throughout the process.

7.2.3 Investigations of neurogenesis in the presence of Smad2CA/3CA

One of the more striking findings from this thesis was the advanced induced neuronal maturity when generated in the presence of Smad2CA/3CA. It will be interesting to look

in further detail as to the mechanism of Smad2CA/3CA actions during neurogenesis. In a similar method to our reprogramming studies, we are preparing time course RNA-seq and ATAC-seq samples to assess global transcriptional and chromatin states throughout the induced neuron transdifferentiation process. The idea will be to assess which pathways are most differentially activated with Smad3CA, to hopefully identify key players involved in maturation of induced neurons. The ATAC-seq analysis should further reveal global chromatin states, revealing how chromatin architecture may correlate to neuronal maturity. The work will also investigate how the resulting neurons perform upon transplantation and their ability to integrate within host neuronal networks, a current challenge in the field. Furthermore, little is known about the underlying mechanisms of the transition from fibroblast to neuron, and how they relate to normal neuron development form ES cells *in-vivo* or *in-vitro*. Recent studies have described roles for Smad3 and TGF-β receptors for neuronal differentiation both *in vitro* and in the developing neural tube. There are strong links between the TGF-β pathway and chromatin remodellers important for neural differentiation. It will be interesting to assess if similar mechanisms are at play for induced neurons, asking if the transdifferentation process shares any similarities to normal developmental contexts.

7.3 Closing remarks

Since the discovery of iPSCs in 2006, an enormous body of work has been dedicated to understand how such an inefficient process occurs. With each discovery, we learn more about the underlying cellular machinery required to orchestrate reprogramming, identify the major barriers, and continuously improve the efficiency of the process. Such advances have undoubtedly enabled a widespread use of the technology, as indicated by the shear number of publications that have utilized iPSCs. Ranging from disease modelling and drug discovery to tissue engineering, the benefits of iPSCs have already been realized. It is an exciting prospect to consider the unlimited potential of such a technology, and we now await the first clinical trials using iPSCs as a source of cells for regenerative therapies.

The fact that iPSCs are widely used and making their way to the clinic, does not mean we no longer need to study the reprogramming process. Recent publications are still reporting reprogramming efficiencies of 1-5%, with exceptional cases claiming higher efficiency, revealing there are still major unidentified roadblocks. While it may not be necessary to increase reprogramming efficiency, the inefficiency of the system itself reveals that there are many characteristics of cell identity conversions we still do not understand. Through studies of reprogramming, we can better manipulate cell identity conversions, as exemplified in this thesis. The generation of various cells and tissues types, hold tremendous promise not only for regenerative therapies, but also to better understand development and manifestation of disease within cells and tissues.

References

Abdollah, S., Macias-Silva, M., Tsukazaki, T., Hayashi, H., Attisano, L., and Wrana, J.L. (1997). TbetaRI phosphorylation of Smad2 on Ser465 and Ser467 is required for Smad2-Smad4 complex formation and signaling. J Biol Chem *272*, 27678-27685.

Acosta, J.C., Snijders, A.P., and Gil, J. (2013). Unbiased characterization of the senescence-associated secretome using SILAC-based quantitative proteomics. Methods Mol Biol *965*, 175-184.

Akiyoshi, S., Inoue, H., Hanai, J., Kusanagi, K., Nemoto, N., Miyazono, K., and Kawabata, M. (1999). c-Ski acts as a transcriptional co-repressor in transforming growth factor-beta signaling through interaction with smads. J Biol Chem *274*, 35269-35277.

Banito, A., Rashid, S.T., Acosta, J.C., Li, S., Pereira, C.F., Geti, I., Pinho, S., Silva, J.C., Azuara, V., Walsh, M.*, et al.* (2009). Senescence impairs successful reprogramming to pluripotent stem cells. Genes Dev *23*, 2134-2139.

Beattie, G.M., Lopez, A.D., Bucay, N., Hinton, A., Firpo, M.T., King, C.C., and Hayek, A. (2005). Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. Stem Cells *23*, 489-495.

Bertero, A., Madrigal, P., Galli, A., Hubner, N.C., Moreno, I., Burks, D., Brown, S., Pedersen, R.A., Gaffney, D., Mendjan, S.*, et al.* (2015). Activin/nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. Genes Dev *29*, 702-717.

Blau, H.M., Chiu, C.P., and Webster, C. (1983). Cytoplasmic activation of human nuclear genes in stable heterocaryons. Cell *32*, 1171-1180.

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell *2*, 151-159.

Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A.*, et al.* (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191-195.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209-1222.

Bussmann, L.H., Schubert, A., Vu Manh, T.P., De Andres, L., Desbordes, S.C., Parra, M., Zimmermann, T., Rapino, F., Rodriguez-Ubreva, J., Ballestar, E.*, et al.* (2009). A robust and highly efficient immune cell reprogramming system. Cell Stem Cell *5*, 554-566.

Carey, B.W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M., and Jaenisch, R. (2009). Reprogramming of murine and human somatic cells using a single polycistronic vector. Proc Natl Acad Sci U S A *106*, 157-162.

Carter, A.C., Davis-Dusenbery, B.N., Koszka, K., Ichida, J.K., and Eggan, K. (2014). Nanog-independent reprogramming to iPSCs with canonical factors. Stem Cell Reports *2*, 119-126.

Cepko, C., and Pear, W. (2001). Overview of the retrovirus transduction system. Curr Protoc Mol Biol *Chapter 9*, Unit9 9.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol *27*, 275-280.

Chantzoura E.; Skylaki, S.M., S.; Kim, S.; Johnsson, A.; Linnarsson, S.; Woltjen, K.; Chambers, I.; Kaji, K. (in press). Reprogramming roadblocks are system-dependent Stem Cell Reports.

Chen, J., Guo, L., Zhang, L., Wu, H., Yang, J., Liu, H., Wang, X., Hu, X., Gu, T., Zhou, Z.*, et al.* (2013a). Vitamin C modulates TET1 function during somatic cell reprogramming. Nat Genet *45*, 1504-1509.

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Wu, Y., Guo, L.*, et al.* (2013b). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat Genet *45*, 34-42.

Chen, T., Yuan, D., Wei, B., Jiang, J., Kang, J., Ling, K., Gu, Y., Li, J., Xiao, L., and Pei, G. (2010). E-cadherin-mediated cell-cell contact is critical for induced pluripotent stem cell generation. Stem Cells *28*, 1315-1325.

Chen, X., Weisberg, E., Fridmacher, V., Watanabe, M., Naco, G., and Whitman, M. (1997). Smad4 and FAST-1 in the assembly of activin-responsive factor. Nature *389*, 85-89.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J.*, et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Cherry, S.R., Biniszkiewicz, D., van Parijs, L., Baltimore, D., and Jaenisch, R. (2000). Retroviral expression in embryonic stem cells and hematopoietic stem cells. Mol Cell Biol *20*, 7419-7426.

Chin, G.S., Liu, W., Peled, Z., Lee, T.Y., Steinbrech, D.S., Hsu, M., and Longaker, M.T. (2001). Differential expression of transforming growth factor-beta receptors I and II and activation of Smad 3 in keloid fibroblasts. Plast Reconstr Surg *108*, 423-429.

Chipuk, J.E., Cornelius, S.C., Pultz, N.J., Jorgensen, J.S., Bonham, M.J., Kim, S.J., and Danielpour, D. (2002). The androgen receptor represses transforming growth factor-beta signaling through interaction with Smad3. J Biol Chem *277*, 1240-1248.

Collado, M., Blasco, M.A., and Serrano, M. (2007). Cellular senescence in cancer and aging. Cell *130*, 223-233.

Conlon, F.L., Barth, K.S., and Robertson, E.J. (1991). A novel retrovirally induced embryonic lethal mutation in the mouse: assessment of the developmental fate of embryonic stem cells homozygous for the 413.d proviral integration. Development *111*, 969-981.

Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B., and Robertson, E.J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development *120*, 1919-1928.

Cordenonsi, M., Dupont, S., Maretto, S., Insinga, A., Imbriano, C., and Piccolo, S. (2003). Links between tumor suppressors: p53 is required for TGF-beta gene responses by cooperating with Smads. Cell *113*, 301-314.

Cui, L., Johkura, K., Yue, F., Ogiwara, N., Okouchi, Y., Asanuma, K., and Sasaki, K. (2004). Spatial distribution and initial changes of SSEA-1 and other cell adhesion-related molecules on mouse embryonic stem cells before and during differentiation. J Histochem Cytochem *52*, 1447-1457.

Dahle, O., Kumar, A., and Kuehn, M.R. (2010). Nodal signaling recruits the histone demethylase Jmjd3 to counteract polycomb-mediated repression at target genes. Sci Signal *3*, ra48.

Davis, H.E., Morgan, J.R., and Yarmush, M.L. (2002). Polybrene increases retrovirus gene transfer efficiency by enhancing receptor-independent virus adsorption on target cell membranes. Biophys Chem *97*, 159-172.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.

Dennler, S., Huet, S., and Gauthier, J.M. (1999). A short amino-acid sequence in MH1 domain is responsible for functional differences between Smad2 and Smad3. Oncogene *18*, 1643-1648.

Derynck, R., and Zhang, Y.E. (2003). Smad-dependent and Smad-independent pathways in TGF-beta family signalling. Nature *425*, 577-584.

Di Stefano, B., Sardina, J.L., van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., and Graf, T. (2014). C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235-239.

Donnelly, M.L., Hughes, L.E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., and Ryan, M.D. (2001). The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. J Gen Virol *82*, 1027-1041.

Dunn, N.R., Vincent, S.D., Oxburgh, L., Robertson, E.J., and Bikoff, E.K. (2004). Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. Development *131*, 1717-1728.

Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naive pluripotency. Science *344*, 1156-1160.

Ebisawa, T., Fukuchi, M., Murakami, G., Chiba, T., Tanaka, K., Imamura, T., and Miyazono, K. (2001). Smurf1 interacts with transforming growth factor-beta type I receptor through Smad7 and induces receptor degradation. J Biol Chem *276*, 12477-12480.

Estaras, C., Akizu, N., Garcia, A., Beltran, S., de la Cruz, X., and Martinez-Balbas, M.A. (2012). Genome-wide analysis reveals that Smad3 and JMJD3 HDM co-activate the neural developmental program. Development *139*, 2681-2691.

Esteban, M.A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., Li, W., Weng, Z., Chen, J., Ni, S.*, et al.* (2010). Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. Cell Stem Cell *6*, 71-79.

Feng, R., Desbordes, S.C., Xie, H., Tillo, E.S., Pixley, F., Stanley, E.R., and Graf, T. (2008). PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. Proc Natl Acad Sci U S A *105*, 6057-6062.

Feng, X.H., Zhang, Y., Wu, R.Y., and Derynck, R. (1998). The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for smad3 in TGF-beta-induced transcriptional activation. Genes Dev *12*, 2153-2163.

Fimia, G.M., Ferreira, R., Passananti, C., and Amati, P. (1995). A polyomavirus enhancer mutant confers ubiquitous high transcriptional efficiency to the SV40 late promoter. Biochem Biophys Res Commun *207*, 339-347.

Freund, A., Patil, C.K., and Campisi, J. (2011). p38MAPK is a novel DNA damage response-independent regulator of the senescence-associated secretory phenotype. EMBO J *30*, 1536-1548.

Funaba, M., and Mathews, L.S. (2000). Identification and characterization of constitutively active Smad2 mutants: evaluation of formation of Smad complex and subcellular distribution. Mol Endocrinol *14*, 1583-1591.

Gaarenstroom, T., and Hill, C.S. (2014). TGF-beta signaling to chromatin: how Smads regulate transcription during self-renewal and differentiation. Semin Cell Dev Biol *32*, 107-118.

Gagliardi, A., Mullin, N.P., Ying Tan, Z., Colby, D., Kousa, A.I., Halbritter, F., Weiss, J.T., Felker, A., Bezstarosti, K., Favaro, R.*, et al.* (2013). A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. EMBO J *32*, 2231-2247.

Galvin, K.E., Travis, E.D., Yee, D., Magnuson, T., and Vivian, J.L. (2010). Nodal Signaling Regulates the Bone Morphogenic Protein Pluripotency Pathway in Mouse Embryonic Stem Cells. Journal of Biological Chemistry *285*, 19747-19756.

Galvin-Burgess, K.E., Travis, E.D., Pierson, K.E., and Vivian, J.L. (2013). TGF-beta-Superfamily Signaling Regulates Embryonic Stem Cell Heterogeneity: Self-Renewal as a Dynamic and Regulated Equilibrium. Stem Cells *31*, 48-58.

Gao, Y., Han, Z., Li, Q., Wu, Y., Shi, X., Ai, Z., Du, J., Li, W., Guo, Z., and Zhang, Y. (2015). Vitamin C induces a pluripotent state in mouse embryonic stem cells by modulating microRNA expression. FEBS J *282*, 685-699.

Garcia-Campmany, L., and Marti, E. (2007). The TGFbeta intracellular effector Smad3 regulates neuronal differentiation and cell fate specification in the developing spinal cord. Development *134*, 65-75.

Garrett-Sinha, L.A., Eberspaecher, H., Seldin, M.F., and de Crombrugghe, B. (1996). A gene for a novel zinc-finger protein expressed in differentiated epithelial cells and transiently in certain mesenchymal cells. J Biol Chem *271*, 31384-31390.

Germain, S., Howell, M., Esslemont, G.M., and Hill, C.S. (2000). Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. Genes Dev *14*, 435-451.

Golipour, A., David, L., Liu, Y., Jayakumaran, G., Hirsch, C.L., Trcka, D., and Wrana, J.L. (2012). A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. Cell Stem Cell *11*, 769-782.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A *89*, 5547-5551.

Graf, T. (2011). Historical origins of transdifferentiation and reprogramming. Cell Stem Cell *9*, 504-516.

Greder, L.V., Gupta, S., Li, S., Abedin, M.J., Sajini, A., Segal, Y., Slack, J.M., and Dutton, J.R. (2012). Analysis of endogenous Oct4 activation during induced pluripotent stem cell reprogramming using an inducible Oct4 lineage label. Stem Cells *30*, 2596-2601.

Greenberg, R.A., O'Hagan, R.C., Deng, H., Xiao, Q., Hann, S.R., Adams, R.R., Lichtsteiner, S., Chin, L., Morin, G.B., and DePinho, R.A. (1999). Telomerase reverse

transcriptase gene is a direct target of c-Myc but is not functionally equivalent in cellular transformation. Oncogene *18*, 1219-1226.

Gronroos, E., Kingston, I.J., Ramachandran, A., Randall, R.A., Vizan, P., and Hill, C.S. (2012). Transforming growth factor beta inhibits bone morphogenetic protein-induced transcription through novel phosphorylated Smad1/5-Smad3 complexes. Mol Cell Biol *32*, 2904-2916.

Gurdon, J.B. (1960). The developmental capacity of nuclei taken from differentiating endoderm cells of Xenopus laevis. J Embryol Exp Morphol *8*, 505-526.

Gurdon, J.B. (1962). Adult frogs derived from the nuclei of single somatic cells. Dev Biol *4*, 256-273.

He, Y., Zhang, H., Yung, A., Villeda, S.A., Jaeger, P.A., Olayiwola, O., Fainberg, N., and Wyss-Coray, T. (2014). ALK5-dependent TGF-beta signaling is a major determinant of late-stage adult neurogenesis. Nat Neurosci *17*, 943-952.

Heyer, J., Escalante-Alcalde, D., Lia, M., Boettinger, E., Edelmann, W., Stewart, C.L., and Kucherlapati, R. (1999). Postgastrulation Smad2-deficient embryos show defects in embryo turning and anterior morphogenesis. Proc Natl Acad Sci U S A *96*, 12595-12600.

Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. Nature *463*, 474-484.

Hoare, M., and Narita, M. (2013). Transmitting senescence to the cell neighbourhood. Nat Cell Biol *15*, 887-889.

Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K., and Yamanaka, S. (2009). Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature *460*, 1132-1135.

Hotta, A., and Ellis, J. (2008). Retroviral vector silencing during iPS cell induction: an epigenetic beacon that signals distinct pluripotent states. J Cell Biochem *105*, 940-948.

Hubackova, S., Krejcikova, K., Bartek, J., and Hodny, Z. (2012). IL1- and TGFbeta-Nox4 signaling, oxidative stress and DNA damage response are shared features of replicative, oncogene-induced, and drug-induced paracrine 'bystander senescence'. Aging (Albany NY) *4*, 932-951.

Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K.*, et al.* (2009a). A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. Cell Stem Cell *5*, 491-503.

Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K.*, et al.* (2009b). A Small-Molecule Inhibitor of Tgf-beta Signaling Replaces Sox2 in Reprogramming by Inducing Nanog. Cell Stem Cell *5*, 491-503.

Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J.B., Lonnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res *21*, 1160-1167.

Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J.B., Lonnerberg, P., and Linnarsson, S. (2012). Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. Nature protocols *7*, 813-828.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lonnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods *11*, 163-166.

Itoh, S., Landstrom, M., Hermansson, A., Itoh, F., Heldin, C.H., Heldin, N.E., and ten Dijke, P. (1998). Transforming growth factor beta1 induces nuclear export of inhibitory Smad7. J Biol Chem *273*, 29195-29201.

James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. (2005). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. Development *132*, 1273-1282.

Jerabek, S., Merino, F., Scholer, H.R., and Cojocaru, V. (2014). OCT4: dynamic DNA binding pioneers stem cell pluripotency. Biochim Biophys Acta *1839*, 138-154.

Jia, S., Wu, D., Xing, C., and Meng, A. (2009). Smad2/3 activities are required for induction and patterning of the neuroectoderm in zebrafish. Dev Biol *333*, 273-284.

Jones, J.B., and Kern, S.E. (2000). Functional mapping of the MH1 DNA-binding domain of DPC4/SMAD4. Nucleic Acids Res *28*, 2363-2368.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature *458*, 771-U112.

Kajimura, S., Seale, P., Kubota, K., Lunsford, E., Frangioni, J.V., Gygi, S.P., and Spiegelman, B.M. (2009). Initiation of myoblast to brown fat switch by a PRDM16-C/EBP-beta transcriptional complex. Nature *460*, 1154-1158.

Kang, Y., Chen, C.R., and Massague, J. (2003). A self-enabling TGFbeta response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. Mol Cell *11*, 915-926.

Karmazinova, M., and Lacinova, L. (2010). Measurement of cellular excitability by whole cell patch clamp technique. Physiol Res *59 Suppl 1*, S1-7.

Karwacki-Neisius, V., Goke, J., Osorno, R., Halbritter, F., Ng, J.H., Weisse, A.Y., Wong, F.C., Gagliardi, A., Mullin, N.P., Festuccia, N.*, et al.* (2013). Reduced Oct4 expression directs a robust pluripotent state with distinct signaling activity and increased enhancer occupancy by Oct4 and Nanog. Cell Stem Cell *12*, 531-545.

Kato, Y., Habas, R., Katsuyama, Y., Naar, A.M., and He, X. (2002). A component of the ARC/Mediator complex required for TGF beta/Nodal signalling. Nature *418*, 641-646.

Kavsak, P., Rasmussen, R.K., Causing, C.G., Bonni, S., Zhu, H., Thomsen, G.H., and Wrana, J.L. (2000). Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation. Mol Cell *6*, 1365-1375.

Kawamura, T., Suzuki, J., Wang, Y.V., Menendez, S., Morera, L.B., Raya, A., Wahl, G.M., and Izpisua Belmonte, J.C. (2009). Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature *460*, 1140-1144.

Kim, D.W., and Lassar, A.B. (2003). Smad-dependent recruitment of a histone deacetylase/Sin3A complex modulates the bone morphogenetic protein-dependent transcriptional repressor activity of Nkx3.2. Mol Cell Biol *23*, 8704-8717.

King, T.J., and Briggs, R. (1955). Changes in the Nuclei of Differentiating Gastrula Cells, as Demonstrated by Nuclear Transplantation. Proc Natl Acad Sci U S A *41*, 321-325.

Kingston, R.E., Chen, C.A., and Rose, J.K. (2003). Calcium phosphate transfection. Curr Protoc Mol Biol *Chapter 9*, Unit 9 1.

Kitamura, T., Koshino, Y., Shibata, F., Oki, T., Nakajima, H., Nosaka, T., and Kumagai, H. (2003). Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. Exp Hematol *31*, 1007-1014.

Kretschmer, A., Moepert, K., Dames, S., Sternberger, M., Kaufmann, J., and Klippel, A. (2003). Differential regulation of TGF-beta signaling through Smad2, Smad3 and Smad4. Oncogene *22*, 6748-6763.

Kriks, S., Shim, J.W., Piao, J., Ganat, Y.M., Wakeman, D.R., Xie, Z., Carrillo-Reid, L., Auyeung, G., Antonacci, C., Buch, A.*, et al.* (2011). Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. Nature *480*, 547-551.

Lakos, G., Takagawa, S., Chen, S.J., Ferreira, A.M., Han, G., Masuda, K., Wang, X.J., DiPietro, L.A., and Varga, J. (2004). Targeted disruption of TGF-beta/Smad3 signaling modulates skin fibrosis in a mouse model of scleroderma. Am J Pathol *165*, 203-217.

Lee, D.S., Shin, J.Y., Tonge, P.D., Puri, M.C., Lee, S., Park, H., Lee, W.C., Hussein, S.M., Bleazard, T., Yun, J.Y.*, et al.* (2014). An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. Nat Commun *5*, 5619.

Levy, L., and Hill, C.S. (2006). Alterations in components of the TGF-beta superfamily signaling pathways in human cancer. Cytokine Growth Factor Rev *17*, 41-58.

Li, H., Collado, M., Villasante, A., Strati, K., Ortega, S., Canamero, M., Blasco, M.A., and Serrano, M. (2009). The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature *460*, 1136-1139.

Li, P., Chen, Y., Meng, X., Kwok, K.Y., Huang, X., Choy, K.W., Wang, C.C., Lan, H., and Yuan, P. (2013). Suppression of malignancy by Smad3 in mouse embryonic stem cell formed teratoma. Stem Cell Rev *9*, 709-720.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., *et al.* (2010a). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell *7*, 51-63.

Li, R.H., Liang, J.L., Ni, S., Zhou, T., Qing, X.B., Li, H.P., He, W.Z., Chen, J.K., Li, F., Zhuang, Q.A., *et al.* (2010b). A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. Cell Stem Cell *7*, 51-63.

Liao, B., Bao, X., Liu, L., Feng, S., Zovoilis, A., Liu, W., Xue, Y., Cai, J., Guo, X., Qin, B., *et al.* (2011). MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. J Biol Chem *286*, 17359-17364.

Lin, T., Ambasudhan, R., Yuan, X., Li, W., Hilcove, S., Abujarour, R., Lin, X., Hahm, H.S., Hao, E., Hayek, A., *et al.* (2009). A chemical platform for improved induction of human iPSCs. Nat Methods *6*, 805-808.

Liu, X., Sun, H., Qi, J., Wang, L., He, S., Liu, J., Feng, C., Chen, C., Li, W., Guo, Y., *et al.* (2013). Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT-MET mechanism for optimal reprogramming. Nat Cell Biol.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lujan, E., Zunder, E.R., Ng, Y.H., Goronzy, I.N., Nolan, G.P., and Wernig, M. (2015). Early reprogramming regulators identified by prospective isolation and mass cytometry. Nature *521*, 352-356.

Luo, K., Stroschein, S.L., Wang, W., Chen, D., Martens, E., Zhou, S., and Zhou, Q. (1999). The Ski oncoprotein interacts with the Smad proteins to repress TGFbeta signaling. Genes Dev *13*, 2196-2206.

Maherali, N., and Hochedlinger, K. (2009). Tgf beta Signal Inhibition Cooperates in the Induction of iPSCs and Replaces Sox2 and cMyc. Current Biology *19*, 1718-1723.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., *et al.* (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55-70.

Malaguti, M., Nistor, P.A., Blin, G., Pegg, A., Zhou, X., and Lowell, S. (2013). Bone morphogenic protein signalling suppresses differentiation of pluripotent cells by maintaining expression of E-Cadherin. Elife *2*, e01197.

Massague, J. (2012). TGFbeta signalling in context. Nat Rev Mol Cell Biol *13*, 616-630.

McTaggart, S., and Al-Rubeai, M. (2002). Retroviral vectors for human gene delivery. Biotechnol Adv *20*, 1-31.

Medici, D., Potenta, S., and Kalluri, R. (2011). Transforming growth factor-beta2 promotes Snail-mediated endothelial-mesenchymal transition through convergence of Smad-dependent and Smad-independent signalling. Biochem J *437*, 515-520.

Miettinen, P.J., Ebner, R., Lopez, A.R., and Derynck, R. (1994). TGF-beta induced transdifferentiation of mammary epithelial cells to mesenchymal cells: involvement of type I receptors. J Cell Biol *127*, 2021-2036.

Miller, J.D., Ganat, Y.M., Kishinevsky, S., Bowman, R.L., Liu, B., Tu, E.Y., Mandal, P.K., Vera, E., Shim, J.W., Kriks, S.*, et al.* (2013). Human iPSC-based modeling of late-onset disease via progerin-induced aging. Cell Stem Cell *13*, 691-705.

Moon, J.H., Yun, W., Kim, J., Hyeon, S., Kang, P.J., Park, G., Kim, A., Oh, S., Whang, K.Y., Kim, D.W.*, et al.* (2013). Reprogramming of mouse fibroblasts into induced pluripotent stem cells with Nanog. Biochem Biophys Res Commun *431*, 444-449.

Morita, S., Kojima, T., and Kitamura, T. (2000). Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene Ther *7*, 1063-1066.

Mu, Y., Gudey, S.K., and Landstrom, M. (2012). Non-Smad signaling pathways. Cell Tissue Res *347*, 11-20.

Mullen, A.C., Orlando, D.A., Newman, J.J., Loven, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., DeKoter, R.P., and Young, R.A. (2011). Master transcription factors determine cell-type-specific responses to TGF-beta signaling. Cell *147*, 565-576.

Mummery, C., Ward-van Oostwaard, D., Doevendans, P., Spijker, R., van den Brink, S., Hassink, R., van der Heyden, M., Opthof, T., Pera, M., de la Riviere, A.B.*, et al.* (2003). Differentiation of human embryonic stem cells to cardiomyocytes: role of coculture with visceral endoderm-like cells. Circulation *107*, 2733-2740.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372-376.

Nomura, M., and Li, E. (1998). Smad2 role in mesoderm formation, left-right patterning and craniofacial development. Nature *393*, 786-790.

O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., and Kaji, K. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317.

Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T., and Yamanaka, S. (2008). Generation of mouse induced pluripotent stem cells without viral vectors. Science *322*, 949-953.

Osakada, F., Ikeda, H., Sasai, Y., and Takahashi, M. (2009). Stepwise differentiation of pluripotent stem cells into retinal cells. Nat Protoc *4*, 811-824.

Oshimori, N., and Fuchs, E. (2012). The harmonies played by TGF-beta in stem cell biology. Cell Stem Cell *11*, 751-764.

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D.R., Yang, T.Q., Citri, A., Sebastiano, V., Marro, S., Sudhof, T.C.*, et al.* (2011). Induction of human neuronal cells by defined transcription factors. Nature *476*, 220-223.

Papapetrou, E.P., Tomishima, M.J., Chambers, S.M., Mica, Y., Reed, E., Menon, J., Tabar, V., Mo, Q., Studer, L., and Sadelain, M. (2009). Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. Proc Natl Acad Sci U S A *106*, 12759-12764.

Papp, B., and Plath, K. (2013). Epigenetics of reprogramming to induced pluripotency. Cell *152*, 1324-1343.

Park, D.H., Hong, S.J., Salinas, R.D., Liu, S.J., Sun, S.W., Sgualdino, J., Testa, G., Matzuk, M.M., Iwamori, N., and Lim, D.A. (2014). Activation of neuronal gene expression by the JMJD3 demethylase is required for postnatal and adult brain neurogenesis. Cell Rep *8*, 1290-1299.

Pera, M.F., Andrade, J., Houssami, S., Reubinoff, B., Trounson, A., Stanley, E.G., Ward-van Oostwaard, D., and Mummery, C. (2004). Regulation of human embryonic stem cell differentiation by BMP-2 and its antagonist noggin. J Cell Sci *117*, 1269-1280.

Pereira, M., Pfisterer, U., Rylander, D., Torper, O., Lau, S., Lundblad, M., Grealish, S., and Parmar, M. (2014). Highly efficient generation of induced neurons from human fibroblasts that survive transplantation into the adult rat brain. Sci Rep *4*, 6330.

Pfisterer, U., Wood, J., Nihlberg, K., Hallgren, O., Bjermer, L., Westergren-Thorsson, G., Lindvall, O., and Parmar, M. (2011). Efficient induction of functional neurons from adult human fibroblasts. Cell Cycle *10*, 3311-3316.

Piek, E., Ju, W.J., Heyer, J., Escalante-Alcalde, D., Stewart, C.L., Weinstein, M., Deng, C., Kucherlapati, R., Bottinger, E.P., and Roberts, A.B. (2001). Functional characterization of transforming growth factor beta signaling in Smad2- and Smad3-deficient fibroblasts. J Biol Chem *276*, 19945-19953.

Piek, E., Moustakas, A., Kurisaki, A., Heldin, C.H., and ten Dijke, P. (1999). TGF-(beta) type I receptor/ALK-5 and Smad proteins mediate epithelial to mesenchymal transdifferentiation in NMuMG breast epithelial cells. J Cell Sci *112 (Pt 24)*, 4557-4568.

Pluck, A., and Klasen, C. (2009). Generation of chimeras by morula aggregation. Methods Mol Biol *561*, 219-229.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J.*, et al.* (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617-1632.

Polo, J.M., and Hochedlinger, K. (2010). When fibroblasts MET iPSCs. Cell Stem Cell *7*, 5-6.

Pouponnot, C., Jayaraman, L., and Massague, J. (1998). Physical and functional interaction of SMADs and p300/CBP. J Biol Chem *273*, 22865-22868.

Qi, X., Li, T.G., Hao, J., Hu, J., Wang, J., Simmons, H., Miura, S., Mishina, Y., and Zhao, G.Q. (2004a). BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. Proc Natl Acad Sci U S A *101*, 6027-6032.

Qi, Y., Gregory, M.A., Li, Z., Brousal, J.P., West, K., and Hann, S.R. (2004b). p19ARF directly and differentially controls the functions of c-Myc independently of p53. Nature *431*, 712-717.

Radzisheuskaya, A., Chia Gle, B., dos Santos, R.L., Theunissen, T.W., Castro, L.F., Nichols, J., and Silva, J.C. (2013). A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nat Cell Biol *15*, 579-590.

Redmer, T., Diecke, S., Grigoryan, T., Quiroga-Negreira, A., Birchmeier, W., and Besser, D. (2011). E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. EMBO Rep *12*, 720-726.

Rizzino, A. (2013). Concise review: The Sox2-Oct4 connection: critical players in a much larger interdependent network integrated at multiple levels. Stem Cells *31*, 1033-1039.

Ruetz, T., and Kaji, K. (2014). Routes to induced pluripotent stem cells. Curr Opin Genet Dev *28*, 38-42.

Saha, K., and Jaenisch, R. (2009). Technical challenges in using human induced pluripotent stem cells to model disease. Cell Stem Cell *5*, 584-595.

Sakaki-Yumoto, M., Liu, J., Ramalho-Santos, M., Yoshida, N., and Derynck, R. (2013). Smad2 is essential for maintenance of the human and mouse primed pluripotent stem cell state. J Biol Chem *288*, 18546-18560.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming. Cell Stem Cell *7*, 64-77.

Sandstrom, R.S., Foret, M.R., Grow, D.A., Haugen, E., Rhodes, C.T., Cardona, A.E., Phelix, C.F., Wang, Y., Berger, M.S., and Lin, C.H. (2014). Epigenetic regulation by chromatin activation mark H3K4me3 in primate progenitor cells within adult neurogenic niche. Sci Rep *4*, 5371.

Schmid, P., Itin, P., Cherry, G., Bi, C., and Cox, D.A. (1998). Enhanced expression of transforming growth factor-beta type I and type II receptors in wound granulation tissue and hypertrophic scar. Am J Pathol *152*, 485-493.

Schneuwly, S., Klemenz, R., and Gehring, W.J. (1987). Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. Nature *325*, 816-818.

Scholer, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N., and Gruss, P. (1989). Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. EMBO J *8*, 2551-2557.

Schwarz, B.A., Bar-Nur, O., Silva, J.C., and Hochedlinger, K. (2014). Nanog is dispensable for the generation of induced pluripotent stem cells. Curr Biol *24*, 347-350.

Seale, P., Kajimura, S., and Spiegelman, B.M. (2009). Transcriptional control of brown adipocyte development and physiological function--of mice and men. Genes Dev *23*, 788-797.

Senturk, S., Mumcuoglu, M., Gursoy-Yuzugullu, O., Cingoz, B., Akcali, K.C., and Ozturk, M. (2010). Transforming growth factor-beta induces senescence in hepatocellular carcinoma cells and inhibits tumor growth. Hepatology *52*, 966-974.

Shao, L., Feng, W., Sun, Y., Bai, H., Liu, J., Currie, C., Kim, J., Gama, R., Wang, Z., Qian, Z*., et al.* (2009). Generation of iPS cells using defined factors linked via the self-cleaving 2A sequences in a single open reading frame. Cell Res *19*, 296-306.

Shi, Y., and Massague, J. (2003a). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell *113*, 685-700.

Shi, Y., Wang, Y.F., Jayaraman, L., Yang, H., Massague, J., and Pavletich, N.P. (1998). Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. Cell *94*, 585-594.

Shi, Y.G., and Massague, J. (2003b). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell *113*, 685-700.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog Is the Gateway to the Pluripotent Ground State. Cell *138*, 722-737.

Sisakhtnezhad, S., and Matin, M.M. (2012). Transdifferentiation: a cell and molecular reprogramming process. Cell Tissue Res *348*, 379-396.

Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells *27*, 543-549.

Souchelnytskyi, S., Tamaki, K., Engstrom, U., Wernstedt, C., ten Dijke, P., and Heldin, C.H. (1997). Phosphorylation of Ser465 and Ser467 in the C terminus of Smad2 mediates interaction with Smad4 and is required for transforming growth factor-beta signaling. J Biol Chem *272*, 28107-28115.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555-568.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230-240.

Staerk, J., Dawlaty, M.M., Gao, Q., Maetzel, D., Hanna, J., Sommer, C.A., Mostoslavsky, G., and Jaenisch, R. (2010). Reprogramming of human peripheral blood cells to induced pluripotent stem cells. Cell Stem Cell *7*, 20-24.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev *64*, 435-459.

Stevens, L.C., and Little, C.C. (1954). Spontaneous Testicular Teratomas in an Inbred Strain of Mice. Proc Natl Acad Sci U S A *40*, 1080-1087.

Suwanabol, P.A., Seedial, S.M., Zhang, F., Shi, X., Si, Y., Liu, B., and Kent, K.C. (2012). TGF-beta and Smad3 modulate PI3K/Akt signaling pathway in vascular smooth muscle cells. Am J Physiol Heart Circ Physiol *302*, H2211-2219.

Suzuki, C., Murakami, G., Fukuchi, M., Shimanuki, T., Shikauchi, Y., Imamura, T., and Miyazono, K. (2002). Smurf1 regulates the inhibitory activity of Smad7 by targeting Smad7 to the plasma membrane. J Biol Chem *277*, 39919-39925.

Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F., and Vignali, D.A. (2004). Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat Biotechnol *22*, 589-594.

Takaesu, N.T., Herbig, E., Zhitomersky, D., O'Connor, M.B., and Newfeld, S.J. (2005). DNA-binding domain mutations in SMAD genes yield dominant-negative proteins or a neomorphic protein that can activate WG target genes in Drosophila. Development *132*, 4883-4894.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Sasaki, A., Yamamoto, M., Nakamura, M., Sutou, K., Osafune, K., and Yamanaka, S. (2014). Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. Nat Commun *5*, 3678.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Tanabe, K., Nakamura, M., Narita, M., Takahashi, K., and Yamanaka, S. (2013). Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. Proc Natl Acad Sci U S A *110*, 12172-12179.

ten Dijke, P., Miyazono, K., and Heldin, C.H. (2000). Signaling inputs converge on nuclear effectors in TGF-beta signaling. Trends Biochem Sci *25*, 64-70.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196-199.

Theunissen, T.W., van Oosten, A.L., Castelo-Branco, G., Hall, J., Smith, A., and Silva, J.C. (2011). Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. Curr Biol *21*, 65-71.

Thiel, G. (2013). How Sox2 maintains neural stem cell identity. Biochem J *450*, e1-2.

Tojo, M., Hamashima, Y., Hanyu, A., Kajimoto, T., Saitoh, M., Miyazono, K., Node, M., and Imamura, T. (2005). The ALK-5 inhibitor A-83-01 inhibits Smad signaling and epithelial-to-mesenchymal transition by transforming growth factor-beta. Cancer Science *96*, 791-800.

Tonge, P.D., Corso, A.J., Monetti, C., Hussein, S.M., Puri, M.C., Michael, I.P., Li, M., Lee, D.S., Mar, J.C., Cloonan, N.*, et al.* (2014). Divergent reprogramming routes lead to alternative stem-cell states. Nature *516*, 192-197.

Tremblay, K.D., Hoodless, P.A., Bikoff, E.K., and Robertson, E.J. (2000). Formation of the definitive endoderm in mouse is a Smad2-dependent process. Development *127*, 3079-3090.

Urlinger, S., Baron, U., Thellmann, M., Hasan, M.T., Bujard, H., and Hillen, W. (2000). Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. Proc Natl Acad Sci U S A *97*, 7963-7968.

Valcourt, U., Kowanetz, M., Niimi, H., Heldin, C.H., and Moustakas, A. (2005). TGF-beta and the Smad signaling pathway support transcriptomic reprogramming during epithelial-mesenchymal cell transition. Mol Biol Cell *16*, 1987-2002.

Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J Cell Sci *118*, 4495-4509.

Varlet, I., Collignon, J., Norris, D.P., and Robertson, E.J. (1997a). Nodal signaling and axis formation in the mouse. Cold Spring Harb Symp Quant Biol *62*, 105-113.

Varlet, I., Collignon, J., and Robertson, E.J. (1997b). nodal expression in the primitive endoderm is required for specification of the anterior axis during mouse gastrulation. Development *124*, 1033-1044.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature *463*, 1035-1041.

Vierbuchen, T., and Wernig, M. (2012). Molecular roadblocks for cellular reprogramming. Mol Cell *47*, 827-838.

Vincent, S.D., Dunn, N.R., Hayashi, S., Norris, D.P., and Robertson, E.J. (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. Genes Dev *17*, 1646-1662.

Vo, N., and Goodman, R.H. (2001). CREB-binding protein and p300 in transcriptional regulation. J Biol Chem *276*, 13505-13508.

Vu, H.N., Ramsey, J.D., and Pack, D.W. (2008). Engineering of a stable retroviral gene delivery vector by directed evolution. Mol Ther *16*, 308-314.

Wakabayashi, Y., Tamiya, T., Takada, I., Fukaya, T., Sugiyama, Y., Inoue, N., Kimura, A., Morita, R., Kashiwagi, I., Takimoto, T.*, et al.* (2011). Histone 3 lysine 9 (H3K9) methyltransferase recruitment to the interleukin-2 (IL-2) promoter is a mechanism of suppression of IL-2 transcription by the transforming growth factor-beta-Smad pathway. J Biol Chem *286*, 35456-35465.

Waldrip, W.R., Bikoff, E.K., Hoodless, P.A., Wrana, J.L., and Robertson, E.J. (1998). Smad2 signaling in extraembryonic tissues determines anterior-posterior polarity of the early mouse embryo. Cell *92*, 797-808.

Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., and Liu, P. (2008). Chromosomal transposition of PiggyBac in mouse embryonic stem cells. Proc Natl Acad Sci U S A *105*, 9290-9295.

Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell *10*, 440-454.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F.*, et al.* (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell *155*, 621-635.

Weinstein, M., Monga, S.P., Liu, Y., Brodie, S.G., Tang, Y., Li, C., Mishra, L., and Deng, C.X. (2001). Smad proteins and hepatocyte growth factor control parallel regulatory pathways that converge on beta1-integrin to promote normal liver development. Mol Cell Biol *21*, 5122-5131.

Wernig, M., Lengner, C.J., Hanna, J., Lodato, M.A., Steine, E., Foreman, R., Staerk, J., Markoulaki, S., and Jaenisch, R. (2008). A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. Nature Biotechnology *26*, 916-924.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318-324.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810-813.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hamalainen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M.*, et al.* (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature *458*, 766-770.

Xi, Q., He, W., Zhang, X.H., Le, H.V., and Massague, J. (2008). Genome-wide impact of the BRG1 SWI/SNF chromatin remodeler on the transforming growth factor beta transcriptional program. J Biol Chem *283*, 1146-1155.

Xi, Q., Wang, Z., Zaromytidou, A.I., Zhang, X.H., Chow-Tsang, L.F., Liu, J.X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V.*, et al.* (2011). A poised chromatin platform for TGF-beta access to master regulators. Cell *147*, 1511-1524.

Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663-676.

Xu, J., Lamouille, S., and Derynck, R. (2009). TGF-beta-induced epithelial to mesenchymal transition. Cell Res *19*, 156-172.

Xu, R.H., Chen, X., Li, D.S., Li, R., Addicks, G.C., Glennon, C., Zwaka, T.P., and Thomson, J.A. (2002). BMP4 initiates human embryonic stem cell differentiation to trophoblast. Nature Biotechnology *20*, 1261-1264.

Xu, R.H., Peck, R.M., Li, D.S., Feng, X., Ludwig, T., and Thomson, J.A. (2005). Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. Nat Methods *2*, 185-190.

Xu, R.H., Sampsell-Barron, T.L., Gu, F., Root, S., Peck, R.M., Pan, G., Yu, J., Antosiewicz-Bourget, J., Tian, S., Stewart, R.*, et al.* (2008). NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. Cell Stem Cell *3*, 196-206.

Yang, X., Letterio, J.J., Lechleider, R.J., Chen, L., Hayman, R., Gu, H., Roberts, A.B., and Deng, C. (1999). Targeted disruption of SMAD3 results in impaired mucosal immunity and diminished T cell responsiveness to TGF-beta. EMBO J *18*, 1280-1291.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.

Ying, Q.L., and Smith, A.G. (2003). Defined conditions for neural commitment and differentiation. Methods Enzymol *365*, 327-341.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci U S A *108*, 1531-1536.

Zawel, L., Dai, J.L., Buckhaults, P., Zhou, S., Kinzler, K.W., Vogelstein, B., and Kern, S.E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. Mol Cell *1*, 611-617.

Zhang, S., Fei, T., Zhang, L., Zhang, R., Chen, F., Ning, Y., Han, Y., Feng, X.H., Meng, A., and Chen, Y.G. (2007). Smad7 antagonizes transforming growth factor beta signaling in the nucleus by interfering with functional Smad-DNA complex formation. Mol Cell Biol *27*, 4488-4499.

Zhang, Y.E. (2009). Non-Smad pathways in TGF-beta signaling. Cell Research *19*, 128-139.

Zheng, Y., Zhao, Y.D., Gibbons, M., Abramova, T., Chu, P.Y., Ash, J.D., Cunningham, J.M., and Skapek, S.X. (2010). Tgfbeta signaling directly induces Arf promoter

remodeling by a mechanism involving Smads 2/3 and p38 MAPK. J Biol Chem *285*, 35654-35664.

Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L., and Kuehn, M.R. (1993). Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. Nature *361*, 543-547.

Zhu, Y., Richardson, J.A., Parada, L.F., and Graff, J.M. (1998). Smad3 mutant mice develop metastatic colorectal cancer. Cell *94*, 703-714.

Zunder, E.R., Lujan, E., Goltsev, Y., Wernig, M., and Nolan, G.P. (2015). A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. Cell Stem Cell *16*, 323-337.